



# **Vergleichende Genomanalyse und mathematische Modellierung von Pathogenitäts- und Regulationenmechanismen bei Bakterien**

Von der Fakultät für Lebenswissenschaften  
der Technischen Universität Carolo-Wilhelmina  
zu Braunschweig

zur Erlangung des Grades einer  
Doktorin der Naturwissenschaften  
(Dr. rer. nat.)

genehmigte  
D i s s e r t a t i o n

von Louisa Roselius  
aus Hamburg, Deutschland

1. Referent: Professor Dr. Dieter Jahn

2. Referent: Professor Dr. Dirk Langemann

Eingereicht am: 29.10.2014

Mündliche Prüfung (Disputation) am: 16.12.2014

Druckjahr 2015

## Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

### Publikationen

Shevchuk O.<sup>1</sup>, **Roselius L.**<sup>1</sup>, Günther G., Klein J., Jahn D., Steinert M., Münch R. (2012).

InFiRe - a novel computational method for the identification of insertion sites in transposon mutagenized bacterial genomes.

Bioinformatics, 28(3):306-310

**Roselius L.**, Langemann D., Müller J., Hense B., Filges S., Jahn D., Münch R. (2014).

Modeling and analysis of a gene-regulatory feed-forward loop with basal expression of the second regulator.

Journal of Theoretical Biology, 363C:290-299

### Tagungsbeiträge

#### Poster

**Roselius L.**, Shevchuk O., Klein J., Günther G., Yarmolinets S., Jahn D., Steinert M., Münch R. (2011)

Computational Prediction of the Transposon Location in Bacterial Genomes  
European Conference on Computational Biology and Annual International Conference on Intelligent System for Molecular Biology, Wien Österreich.

---

<sup>1</sup> geteilte Erst-Autorenschaft

**Vortrag**

**Roselius L.**, Langemann D., Müller J., Hense B., Jahn D., Münch R. (2014)  
Modeling and analysis of a gene-regulatory feed-forward loop with basal expression of the second regulator  
ECMTB, 9th European Conference on Mathematical and Theoretical Biology, Göteborg Schweden.





**Für meine Eltern**





# Inhaltsverzeichnis

<b>Verzeichnisse</b>	<b>VI</b>
Glossar . . . . .	VII
Variablenverzeichnis . . . . .	X
Abbildungsverzeichnis . . . . .	XIII
Tabellenverzeichnis . . . . .	XVII
 <b>1. Zusammenfassung</b>	 <b>1</b>
 <b>2. Einführung</b>	 <b>7</b>
2.1. Biomathematik . . . . .	8
2.2. Bioinformatik . . . . .	9
2.2.1. Sequenzdatenbanken . . . . .	10
2.3. “Omics”-Technologien . . . . .	11
2.4. Systembiologie . . . . .	13
2.5. Programmierung-Software . . . . .	15
2.5.1. Matlab . . . . .	15
2.5.2. R . . . . .	16
2.6. Genregulation . . . . .	18
2.6.1. Genregulatorische Netzwerke . . . . .	20
2.7. Modellierung von Genregulationen über Differentialgleichun- gen . . . . .	23
2.8. Ziel der Arbeit . . . . .	26

<b>I. <i>InFiRe</i> - Identifikation von Transposon-Insertionsstellen in bakteriellen Genomen</b>	<b>29</b>
<b>3. Einleitung</b>	<b>31</b>
3.1. Transposon . . . . .	32
3.1.1. Transposon-Klassen . . . . .	32
3.1.2. Transposon-Mutagenese . . . . .	33
3.2. Restriktionsenzyme . . . . .	34
3.3. Southern Blot . . . . .	36
<b>4. Aufbau und Methoden des <i>InFiRe</i>-Algorithmus</b>	<b>37</b>
4.1. Fragmentlängenverteilung . . . . .	38
4.2. Fragmentsuche in bakterieller DNA nach einem Restriktionsverdau . . . . .	40
4.3. <i>InFiRe</i> -Algorithmus . . . . .	43
4.4. Arbeitsablauf des <i>InFiRe</i> -Algorithmus . . . . .	45
4.5. <i>R</i> -Pakete . . . . .	48
4.5.1. <i>Digestion</i> -Paket . . . . .	48
4.5.2. <i>InFiRe</i> -Paket . . . . .	49
<b>5. Ergebnisse und Diskussion</b>	<b>50</b>
5.1. Statistik . . . . .	50
5.1.1. Globale Statistik . . . . .	51
5.1.2. Explizite Statistik . . . . .	54
5.1.3. Vergleich der Statistiken . . . . .	59
5.2. Analyse innerhalb des Webinterface . . . . .	60
5.3. Fallstudie . . . . .	64
5.4. <i>InFiRe</i> im Vergleich zu bisherigen Verfahren . . . . .	67
5.5. Ausblick . . . . .	69

<b>II. Modellierung und Analyse von einer genregulatorischen Feedforward-Schleife mit basaler Expression des zweiten Regulators</b>	<b>71</b>
<b>6. Einleitung</b>	<b>73</b>
6.1. <i>Pseudomonas aeruginosa</i> . . . . .	74
6.2. Denitrifikation in <i>P. aeruginosa</i> . . . . .	75
6.2.1. Regulation der Nitratatmung in <i>P. aeruginosa</i> . . . . .	77
6.3. Feedforward-Schleife . . . . .	80
6.3.1. Filterung . . . . .	82
6.3.2. Basale Expression des zweiten Regulators . . . . .	84
<b>7. Modellierung der Feedforward-Schleife mit basaler Expression</b>	<b>86</b>
7.1. Signal $s_x$ . . . . .	87
7.2. Differentialgleichungsmodell . . . . .	87
7.3. Nutzenbilanz-Funktion $C(k)$ . . . . .	90
7.4. Voraussetzungen . . . . .	93
<b>8. Ergebnisse und Diskussion</b>	<b>98</b>
8.1. Lineares Modell . . . . .	98
8.1.1. Nutzenbilanz bei Signalwechsel . . . . .	99
8.1.2. Maximum der basalen Expression . . . . .	102
8.1.3. Simulation mit linearem Modell . . . . .	104
8.2. Nichtlineares Modell . . . . .	107
8.2.1. Nichtlineare Produktionsfunktion . . . . .	108
8.2.2. Maximum der basalen Expression . . . . .	109
8.2.3. Grenzbereiche . . . . .	113
8.2.4. Simulation mit dem nichtlinearen Modells . . . . .	116
8.3. Modell mit Speicher . . . . .	118
8.3.1. Notwendigkeit der Erweiterung . . . . .	118

8.3.2.	Energiespeicher . . . . .	119
8.3.3.	Lineares Modell mit Speicher . . . . .	124
8.3.4.	Nichtlineares Modell mit Speicher . . . . .	128
8.3.5.	Vergleich des linearen und des nichtlinearen Modells mit Speicher . . . . .	139
8.3.6.	Aufbau des zweiten Regulators nach Signalwechsel .	140
8.4.	Evolutionärer Vorteil der Modelle . . . . .	143
8.5.	<i>Pseudomonas aeruginosa</i> . . . . .	145
8.6.	Ausblick . . . . .	146

### **III. Modellierung einer gepaarten positiven und negativen Feedback-Schleife von RovA in pathogenem *Yersinia pseudotuberculosis* 147**

#### **9. Einleitung 149**

9.1.	<i>Yersinia pseudotuberculosis</i> . . . . .	150
9.1.1.	Infektion . . . . .	150
9.1.2.	RovA . . . . .	151
9.2.	Feedback-Schleife . . . . .	156
9.2.1.	Negative Feedback-Schleife . . . . .	156
9.2.2.	Positive Feedback-Schleife . . . . .	157
9.3.	Bistabilität . . . . .	158
9.3.1.	Biologisches bistabiles Verhalten . . . . .	159
9.3.2.	Potential . . . . .	160
9.4.	Hysterese . . . . .	161
9.5.	Phasendiagramme . . . . .	162

#### **10. Modellierung der temperaturabhängigen RovA-Bistabilität 164**

10.1.	RovA-Produktion als Differentialgleichung . . . . .	165
10.2.	Temperaturabhängige Abbaurate von RovA . . . . .	168

10.3. Temperaturabhängige DNA-Bindungskonstanten von RovA .	170
10.4. Temperaturwechsel Experiment . . . . .	176
10.4.1. Statistische Berechnung der Expressionsraten . . . .	178
10.5. Gesamt-Modell . . . . .	180
<b>11. Ergebnisse und Diskussion</b>	<b>181</b>
11.1. Charakterisierung der RovA-Promotoren . . . . .	181
11.2. Analyse der Bistabilität . . . . .	183
11.2.1. Einfluss der basalen Expressionsrate . . . . .	188
11.2.2. Einfluss der maximalen Expressionsrate . . . . .	190
11.2.3. Einfluss der Degradation . . . . .	192
11.3. Erstellen der <i>in silico</i> -Mutante $\Delta$ iRovA . . . . .	194
11.4. Analyse der Bistabilität der <i>in silico</i> -Mutante $\Delta$ iRovA . . . .	198
11.4.1. Einfluss der basalen Expressionsrate . . . . .	198
11.4.2. Einfluss der maximalen Expressionsrate . . . . .	200
11.4.3. Einfluss der Degradation . . . . .	201
11.4.4. Potential . . . . .	203
11.5. Modell und natürliches Verhalten von RovA im Vergleich . .	204
11.6. Ausblick . . . . .	207
<b>Literaturverzeichnis</b>	<b>209</b>
<b>Index</b>	<b>224</b>
<b>Danksagung</b>	<b>229</b>



# Glossar

## Teil I

### **InFiRe - Identifikation von Insertionsstellen von Transposonmutierten bakteriellen Genomen**

ATP	Adenosintriphosphat
BLAST	Basic Local Alignment Search Tool - Programm zur Analyse von Sequenzen
bp	Basenpaare
cDNA	Complementary DNA (deutsch: komplementäre DNS)
DNA	Deoxyribonucleic acid (deutsch: Desoxyribonukleinsäure)
<i>InFiRe</i>	<u>I</u> n <u>s</u> ertion <u>F</u> inder via <u>R</u> estriction digestion
IUPAC	International Union of Pure and Applied Chemistry
PCR	Polymerase Chain Reaction (deutsch: Polymerase-Kettenreaktion)
R	Freie Programmiersprache ("The R Project for Statistical Computing")
RNA	Ribonucleic acid (deutsch: Ribonukleinsäure)

## Teil II

### **Modellierung und Analyse von einer genregulatorischen Feedforward-Schleife mit basaler Expression des zweiten Regulators**

FFL	Feedforward-Schleife, englisch: feedforward loop
$\text{FFL}_k$	Feedforward-Schleife mit basaler Expression $k$ im zweiten Regulator $y$ .
Mbp	Mega-Basenpaare (1Mbp=1000 bp)

## Teil III

### **Modellierung einer gepaarten positiven und negativen Feedback-Schleife von RovA in pathogenem *Yersinia* *pseudotuberculosis***

$\Delta\text{iRovA}$	<i>In silico</i> -Mutante von <i>Y. pseudotuberculosis</i> , bei dem das RovA-System keinen negativen Feedback-Loop besitzt.
FBL	Feedback-Schleife, englisch feedback loop
<i>inv</i>	Invasin-Gen
DNA	Deoxyribonucleic acid (deutsch: Desoxyribonukleinsäure)
LB	Komplexes Nährmedium für Bakterienkultur, englisch: lysogeny broth
NFL	Negative Feedback-Schleife



PFL	Positive Feedback-Schleife
RovA	Globaler Virulenzregulator in <i>Y. pseudotuberculosis</i>
RSS	Abweichungsquadrate, englisch: Residual Sum of Square

# Variablenverzeichnis

## Teil I

### InFiRe - Identifikation von Insertionsstellen von Transposonmutierten bakteriellen Genomen

$e$	Fehlerwahrscheinlichkeit der Anzahl an berechneten Restriktionsverdaue
$F_k$	Durch <i>InFiRe</i> berechnet Fragmentlängen (in [bp])
$\ell$	Länge der Erkennungssequenz von Restriktionsenzymen
$L(z^s)$	Zuordnung von dem Fragment $z^s$ zu dessen Länge
$L^{-1}$	Zuordnung einer Fragmentlänge zu einer Menge $M$ an Fragmenten
$m$	Durchschnittliche Fragmentlänge nach Restriktionsreaktion
$M$	Menge an Fragmenten mit einer Länge
$n$	Basenanzahl eines Genoms
$p$	Betrag an verschiedenen Nukleinbasen (meist $p = 4$ )
$P_1$	Theoretische Dichtefunktion von Fragmentgrößen ohne Ableseungenauigkeit
$P_2$	Theoretische Dichtefunktion von Fragmentgrößen mit Ableseungenauigkeit

$Q_i$	Alle Fragmente die auf einem Fragmentlängen-Intervall liegen
$S$	Die durchschnittliche Anzahl an Schnitten eines Restriktionsenzym innerhalb eines Genoms
$U$	Übereinstimmung der gemessenen Fragmentlängen und der berechneten Fragmentlängen
$x$	Fragmentlänge in bp
$x_i^{\max}$	Fragmentlänge mit maximal vielen zugeordneten Fragmenten
$z$	Anzahl an möglichen Genompositionen der Transposoninsertion
$z^s$	Neue kleinere genomische Fragmente nach einem Restriktionsverdau

## Teil II

# Modellierung und Analyse von einer genregulatorischen Feedforward-Schleife mit basaler Expression des zweiten Regulators

$a_k$	Energieverteiler im FFL-Modell mit Speicher abhängig von der basalen Expression $k$
$b_k$	Energiespeicher im FFL-Modell mit Speicher abhängig von der basalen Expression $k$
$C(k)$	Nutzenbilanz-Funktion
$\eta_y$	Kostenfaktor mit denen die Produktion von $y$ Kosten verursacht
$\eta_z$	Kostenfaktor mit denen die Produktion von $z$ Kosten verursacht
$g_k(x)$	Produktionsfunktion des zweiten Regulators $y$
$h$	Hill-Koeffizienten
$H_{xy}(t)$	Sättigungsfunktionen von $y$ abhängig von $x$

---

$j_1$	Energiefluss in den Speicher aus dem Verteiler
$j_2$	Energiefluss aus dem Speicher in den Verteiler
$j_{\text{in}}$	Rate des Energieflusses in den Verteiler $a$
$j_{\text{out}}$	Rate des Energieflusses in den Grundmetabolismus
$j_w$	Wachstumsrate
$j_y$	Energiebedarf zur Produktion des zweiten Regulators $y_k$
$j_z$	Energiebedarf zur Produktion der Zielvariable $z_k$
$k$	Basale Expression des zweiten Regulators $y_k$ mit $k \in [0, 1]$
$K_{xy}$	Sättigungskonstante bei Funktion $H_{xy}(t)$
$L$	Linearer Operator mit $\dot{u} = f - u$ ist $L[f] = u$ .
$m(x, y_k)$	Produktionsfunktion der Zielvariable $z_k$
$O$	Landau-Symbol (asymptotische obere Schranke)
$s_x$	1. externes Signal (z.B.: $O_2$ )
$s_y$	2. externes Signal (z.B.: $NO_3^-$ )
$T$	Periodenlänge (Zeit)
$T_i$	Zeitpunkte des Signalwechsel
$\tau$	Zeitintervall
$w_k$	Fitness abhängig von der basalen Expression $k$
$\dot{w}_k$	Wachstum abhängig von der basalen Expression $k$
$x(t)$	1. Regulator der FFL
$y_k(t)$	2. Regulator der FFL
$z_k(t)$	Zielvariable der FFL

## Teil III

# Modellierung einer gepaarten positiven und negativen Feedback-Schleife von RovA in pathogenem *Yersinia pseudotuberculosis*

$A_{\text{RovA}}$	Aktivität der RovA-Promotoren
$\alpha_0$	Basale Expressionsrate von RovA
$\alpha$	Maximale Expressionsrate von RovA
$h$	Hill-Koeffizienten
$h_a$	Hill-Koeffizienten des aktivierenden Promotors
$h_r$	Hill-Koeffizienten des reprimierenden Promotors
$\delta(T)$	Abbaurrate von RovA abhängig von der Temperatur $T$
$k_a(T)$	Bindekonstante an den aktivierenden Promotor $P_1$ abhängig von der Temperatur $T$
$k_r(T)$	Bindekonstante an den reprimierenden Promotor $P_2$ abhängig von der Temperatur $T$
$P_1$	Starker aktivierender Promotor im RovA-Operon
$P_2$	Schwacher reprimierender Promotor im RovA-Operon
$r(t)$	Konzentration von RovA
$\dot{r}$	Veränderung der Konzentration von RovA
$T$	Temperatur in °C

# Abbildungsverzeichnis

2.1.	Aufbau eines prokaryotischen Operons . . . . .	20
2.2.	Drei-Gen-Netzwerkmotive in der Genregulation . . . . .	22
3.1.	Schnittarten von Restriktionsenzymen . . . . .	35
4.1.	Exponentialverteilung der Fragmentlängen nach einem Restriktionsverdau und theoretische Exponential Verteilung .	38
4.2.	Häufigkeitsverteilung der Fragmentlängen mit Fragment-Suche . . . . .	42
4.3.	Beispiel der Fragmentüberlagerungen . . . . .	44
4.4.	Arbeitsablauf zur Detektion von Transposoninsertions-Stellen mittels <i>InFiRe</i> . . . . .	46
5.1.	Die theoretische Anzahl an benötigten Restriktionsreaktionen . . . . .	54
5.2.	Beispiel für die Fragmentlängen-Verteilung . . . . .	56
5.3.	Stufen der Statistik des Webinterfaces . . . . .	57
5.4.	Resultat der <i>InFiRe</i> -Statistik . . . . .	58
5.5.	Stufen der Analyse des Webinterfaces . . . . .	62
5.6.	Southern Blot Hybridisierung nach Restriktionsreaktionen von <i>L. pneumophila</i> Corby . . . . .	65
5.7.	Resultat des <i>InFiRe</i> -Algorithmus . . . . .	66
6.1.	Regulatorisches Netzwerk der Denitrifikation von <i>Pseudomonas aeruginosa</i> . . . . .	78

6.2.	Typ-1 kohärente Feedforward-Schleife . . . . .	81
6.3.	Filterung von Pulsen durch eine FFL . . . . .	83
8.1.	Die Nutzenbilanz direkt nach dem ersten Wechsel des Signals innerhalb des linearen Modells . . . . .	101
8.2.	Simulationsergebnis des linearen Modells . . . . .	106
8.3.	Produktionsfunktion des zweiten Regulators $y_k$ . . . . .	109
8.4.	Die Nutzenbilanz direkt nach dem ersten Wechsel des Signals innerhalb des nichtlinearen Modells . . . . .	115
8.5.	Simulationsergebnis des nichtlinearen Modells . . . . .	117
8.6.	Das Modell der FFL mit Speicher . . . . .	120
8.7.	Die Fitness $w_k$ des linearen Modells mit Energiespeicher für variierende basale Expressionen direkt nach einem Signalwechsel $T_1$ . . . . .	124
8.8.	Simulations-Ergebnisse des linearen Modells mit Speicher	126
8.9.	Optimale basale Expression des linearen Speichermodells .	127
8.10.	Die Fitness $w_k$ direkt nach dem ersten Wechsel des Signals innerhalb des nichtlinearen Modells mit Speicher . . . . .	129
8.11.	Simulationsergebnis des linearen Modells mit Speicher . .	133
8.12.	Optimale Fitness des nichtlinearen Speichermodells . . . .	134
8.13.	Optimale basale Expression des nichtlinearen Speichermodells . . . . .	137
8.14.	Optimale basale Expression des linearen und nichtlinearen Speichermodells . . . . .	139
8.15.	Abbau und erneuter Aufbau des zweiten Regulators . . . .	141
8.16.	Benötigte Zeit zum Aufbau des zweiten Regulators nicht-normiert und normiert . . . . .	142
9.1.	Operon von RovA bei 25°C . . . . .	152
9.2.	Operon von RovA bei 37°C . . . . .	154

9.3.	Regulationskreislauf von RovA abhängig von der Temperatur . . . . .	155
9.4.	Potential eines bistabilen Verhalten eines Systems . . . . .	161
9.5.	Phasendiagramm der Bistabilität des RovA-Modells zu der Temperatur und der maximalen Expressionsrate $\alpha$ . . . . .	162
10.1.	RovA-Operon bei 25°C und 37°C . . . . .	166
10.2.	Abbaurationen von RovA in Abhängigkeit von der Temperatur	169
10.3.	Electrophoretic Mobility Shift Assay für beide Promotoren der RovA-Rproduktion bei zwei Temperaturen . . . . .	172
10.4.	Temperaturabhängigkeit der Bindekonstanten $k_a(T)$ und $k_r(T)$	175
10.5.	Temperaturwechsel Experiment mit einem RovA-Promotor-Reporter-System . . . . .	176
10.6.	Temperaturshift Experiment mit Western-Blot . . . . .	177
11.1.	Promotor-Aktivität und Produktionsfunktion von RovA . .	182
11.2.	Reaktionszeit der RovA-Anpassung von höherer und niedriger Temperatur . . . . .	185
11.3.	Gleichgewichtszustand der RovA-Konzentration abhängig von der Starttemperatur . . . . .	187
11.4.	Phasendiagramm des bistabilen Verhaltens des RovA-Systems abhängig von der Temperatur und der basalen Expressionsrate $\alpha_0$ . . . . .	189
11.5.	Phasendiagramm des bistabilen Verhaltens des RovA-Systems abhängig von der Temperatur und der maximalen Expressionsrate $\alpha$ . . . . .	191
11.6.	Phasendiagramm des bistabilen Verhaltens des RovA-Systems abhängig von der Temperatur und der Abbauration $\delta(T)$ . . .	193
11.7.	Reaktionszeit der RovA-Anpassung von höherer und niedriger Temperatur des Mutationssystems . . . . .	195



---

11.8.	Gleichgewichtszustand der RovA-Konzentration abhängig von der Starttemperatur der $\Delta$ iRovA-Mutante und des Wild-typs . . . . .	197
11.9.	Phasendiagramm des bistabilen Verhaltens des RovA-Mutations-systems abhängig von der Temperatur und der basalen Ex-pressionsrate $\alpha_0$ . . . . .	199
11.10.	Phasendiagramm des bistabilen Verhaltens des RovA-Mutations-systems abhängig von der Temperatur und der maximalen Expressionsrate $\alpha$ . . . . .	201
11.11.	Phasendiagramm des bistabilen Verhaltens des RovA-Mutations-systems abhängig von der Temperatur und der Abbaurrate $\delta(T)$ . . . . .	202
11.12.	Skalarpotential des RovA-Modells abhängig von der Tem-peratur und der RovA-Konzentration . . . . .	204

# Tabellenverzeichnis

2.1.	Primäre Sequenzdatenbanken . . . . .	11
3.1.	Typen von Restriktionsenzymen . . . . .	35
5.1.	Restriktionsreaktions-Listen . . . . .	60
5.2.	Fragmentlängen nach Restriktionsreaktion . . . . .	66
6.1.	Elektronenakzeptoren, Regulationsenzyme und Gene der Nitratatmung . . . . .	79
10.1.	Involvierte Parameter in die Modellierung der RovA Pro- duktion . . . . .	167
10.2.	Abbauraten von RovA für verschiedene Temperaturen . . .	169
10.3.	Berechnung der Bindekonstanten für den aktivierenden und reprimierenden RovA-Promotor . . . . .	173
10.4.	Parameter des RovA-Modells . . . . .	180





# 1. Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der Entwicklung bioinformatischer Methoden und mathematischer Modellierung im Bereich der Mikrobiologie. Sie ist in drei Teile gegliedert.

Der erste Teil beschäftigt sich mit der Lokalisation von Transposons mit Hilfe vergleichender Genomanalyse. Die Position inserierter Transposons in bakteriellen Genomen zu analysieren, ist eine wichtige Möglichkeit Genfunktionen zu bestimmen. Nach einer Transposonmutation wird normalerweise der Genotyp durch Sequenzierungsverfahren analysiert. Ein Algorithmus mit dem Namen *InFiRe*, Insertion Finder via Restriction digestion, wird hierfür entwickelt, welcher Transposoninsertionen innerhalb von Prokaryotengenomen über individuelle Verteilungen von Restriktionsenzymststellen lokalisieren kann. Nach einem Restriktionsverdau und einer anschließenden Southern-Blot-Hybridisierung können mehrere mögliche Positionen der Insertion innerhalb des bakteriellen Genoms zugeordnet werden. Diese Beobachtung kann mittels wiederholten Restriktionsverdauen verfeinert und zu einer genaueren Kartierung der Transposons genutzt werden. Eine anschließende Sequenzierung ist nach dieser Methode nicht mehr notwendig, kann das Ergebnis aber bestätigen. Durch *InFiRe* kann daher eine schnelle und kostengünstige Bestimmung der Transposoninsertionsstelle innerhalb des bakteriellen Genom erfolgen. Auch unter Bedingungen, bei denen bisherige Verfahren eine Lokalisierung von Transposoninsertionen erschwert

haben, kann diese Methode funktionieren. Der Algorithmus ist in Form einer Software im Internet frei zur Verfügung gestellt.

Im zweiten Teil der Arbeit wird mit Hilfe mathematischer Modellierung ein Regulations-Motiv untersucht. Bakterien müssen sich an wechselnde Umgebungsbedingungen anpassen. Mittels genregulatorischer Netzwerke kann diese Anpassung gesteuert werden. Diese Netzwerke setzen sich aus externen Signalen, internen Regulatoren und Zielproteinen, bzw. Zielgenen zusammen. Mittels Modellierung von Differentialgleichungen können die Netzwerke mathematisch dargestellt, analysiert und ausgewertet werden. Ein gesamtes Netzwerk kann verschiedene Subregulationen beinhalten, die sogenannten Regulations-Motive. Innerhalb dieser Arbeit wird eine sogenannte Feedforward-Schleife (FFL - engl. feedforward loop) mit einem dauerhaft basal exprimierten internen Regulator modelliert. Die basale Produktion dieses Regulators kostet dabei innerhalb der Regulation Energie. Mittels einer Kosten-Nutzen-Optimierung wird dieses Regulations-Motiv unter bestimmten wechselnden Umgebungen analysiert. Diese Art der Regulation kann in *Pseudomonas aeruginosa* innerhalb der anaeroben Anpassung beobachtet werden. Durch die Modellierung der FFL kann das Verhalten von *P. aeruginosa* bei wechselnden Sauerstoffbedingungen genauer betrachtet und vorherige Beobachtungen bestätigt werden.

Auch der dritte Teil der Arbeit beschäftigt sich mit der mathematischen Modellierung eines genregulatorischen Netzwerke-Motivs. Hierbei handelt es sich um das häufig vorkommende Motiv einer Feedback-Schleife (FBL - engl. feedback loop). Im konkreten Fall ist die Regulation innerhalb der Initialisierung einer Infektion von *Yersinia pseudotuberculosis*, gesteuert durch den globalen Virulenzregulator RovA, untersucht. RovA wird über eine positive und eine negative FBL autoreguliert und von einer temperatursensitiven Protease abgebaut. Die Regulation von RovA ist eines der ersten Beispiele für temperaturabhängige Bistabilität in Bakterien. Die Bistabilität des

RovA-Steuerungssysteme ist exakt auf den Bereich zwischen Umgebungs- und Wirtstemperatur angepasst. Das Modell und dessen Parameter ist mit Experimenten verifiziert und validiert. Das Zusammenspiel der einzelnen Parameter auf die Bistabilität, das zeitliche Verhalten der Regulation und der Einfluss der negativen FBL sind untersucht worden.

# Summary

This thesis deals with the development of bioinformatical methods and mathematical modelling in the microbiology domain. The work is divided into three parts.

The first part utilises comparative genome analysis to localise transposon insertion sites in bacteria genomes. The position of these transposon insertion sites is an important tool to determine gene functions. After a transposon mutation, the genotype of a bacterial genome has to be identified. Usually the genotype is determined by DNA sequencing. A new algorithm called *InFiRe*, Insertion Finder via Restriction digestion, which has been developed during this thesis, detects transposon insertion sites in prokaryotic genomes by using the characteristic distribution of restriction cuts. After restriction digestion and a subsequent Southern-Blot hybridisation, the transposon insertion site can be associated within several possible positions in the bacterial genome. The observation can be refined by repeated restriction digestion and *InFiRe*. This may lead to a more accurate mapping of the insertion. Sequencing is not required with this method, but can confirm the result. *InFiRe* is a fast and cheap alternative to previous methods of transposon insertion site detection. It works even under circumstances under which former techniques do not. The algorithm is publicly available in form of a software.

The second part investigates a regulation motif with mathematical model-



ling. Bacteria have to adapt to changing environments. The adaptation is regulated with the help of regulatory networks, which were composed of external signals, internal regulators, and targets. These networks can be mathematically modeled, analyzed and evaluated by differential equation systems. Subsets of huge regulation systems are called regulatory motives. In this work one particular regulatory loop, the feedforward loop (FFL) with a basal expression of one of the regulators, is modelled. The basal expression of one regulator is on one hand energetically expensive. This kind of regulation is studied for fluctuating environments with a cost–benefit analysis. This regulation could be observed in the regulation of the denitrification in *Pseudomonas aeruginosa*. The behavior of *P. aeruginosa* at switching oxygen conditions could be considered in more detail and previous observations could be confirmed.

Also the third part of the work concerned with mathematical modelling of gene regulatory network-motive, the frequently occurred feedback loop (FBL). This kind of regulation could be observed for example at the initialization of the infection of *Yersinia pseudotuberculosis*, which is controlled by the global virulence factor RovA. The production of RovA is regulated by one positive and one negative autoregulative FBL and is degraded by a temperature-sensitive protease. The regulation of RovA is a first example of temperature dependent bistable regulation in bacteria. The bistability of the RovA concentration is adapted exactly to the temperature range between host and external environment. The derived model was verified and refined by experimental observations. One focus of this study has been on the influence of the parameters on the bistability, the required time of adaptation, and the effect of the negative FBL.



## 2. Einführung

Experimentelle Beobachtungen können mit biomathematischer Modellierung oder bioinformatischer Analysen untersucht werden. Aus den Beobachtungen werden Thesen formuliert, welche durch Modelle oder Analysen konkretisiert oder verifiziert und durch erneute Experimente bestätigt werden. Informatik und Mathematik sind bei der Auswertung biologischer Daten von großem Nutzen, um Experimente sowohl zeit- als auch kostenoptimiert zu beschreiben und analysieren.

Die hier angewandte mathematische Modellierung von biologischen Vorgängen ist ein Bestandteil der Systembiologie. Innerhalb der Systembiologie sollen biologische Prozesse, Interaktionen und Regulationen gesamtheitlich über die verschiedenen Ebenen eines Systems dargestellt werden. Diese Ebenen können Systemeigenschaften vom Genom bis zum Proteom oder Eigenschaften von Zellen, Geweben und Organismen sein. Die Systembiologie ist eine auf der Biologie basierende, interdisziplinäre Wissenschaft, welche die Biologie mit der Informatik und der Mathematik verknüpft. Diese Arbeit gliedert sich in drei Teile, welche sich in den Bereichen Bioinformatik, Systembiologie und biomathematische Modellierung befinden. Alle drei Teile nutzen experimentelle Beobachtungen, welche durch mathematische Modelle oder bioinformatische Analysen verfeinert oder konkretisiert werden.

## 2.1. Biomathematik

Die Biomathematik beinhaltet die theoretische Biologie und die Biostatistik und ist eine interdisziplinäre Naturwissenschaft. Sie ist Teilgebiet der angewandten Mathematik und nutzt zum Beispiel innerhalb der theoretischen Biologie mathematische Modelle zur Beschreibung von biologischen Prozessen oder kann mittels der Biostatistik biologische Daten statistisch auswerten. Die Biomathematik kombiniert Themen wie Biologie, Evolutionstheorie, Biotechnologie, Medizin, Pharmazie und auch Analysis, Statistik, Algebra, Dynamische Systeme, Graphentheorie, Diskrete Mathematik und vieles mehr.

Die theoretische Biologie hatte ihren Ursprung am Ende des 19., bzw. Anfang des 20. Jahrhundert. 1901 wurde das erste Buch dieses Themas veröffentlicht, *Einleitung in die theoretische Biologie* von Johannes Reinke (Reinke, 1901). Zu der theoretische Biologie gehören mathematische Modellierung von biologischen Prozessen mittels dynamischer Systeme. Die ersten beschriebenen biomathematischen Systeme waren Populationssysteme, im Speziellen in den 1920er Jahren das Lotka-Volterra-Modell, nach Vito Volterra und Alfred Lotka. Dieses beschreibt die Population von Räuber- und Beutetieren mittels eines Differentialgleichungssystem (Lotka, 1922, Volterra, 1931). Auch viele andere biologische Prozesse können mittels dynamischer Systeme modelliert werden. Dazu gehören neuronale Netzwerke (Latham et al., 2000), Epidemiologie (Oliva and Sallum, 1996), Metabolische Netzwerke (Steuer et al., 2006), zelluläre Prozesse (Villani and Serra, 2013) oder auch Genregulationen.

Die Biostatistik ist ein Bereich der Biomathematik und hat sich erst im letzten Viertel des 20. Jahrhunderts weit verbreitet. Durch die immer größer werdenden Datensätze, zum Beispiel durch sequenzierte Genome oder durch Daten von Hochdurchsatzmethoden, wurde die Biostatistik ein essen-

tieller Bestandteil innerhalb der Biologie. Die statistische Auswertung in der Biologie beinhaltet die deskriptive und die analytische Statistik. So können zum Beispiel durch die Biostatistik Microarrays ausgewertet werden (Wettenhall and Smyth, 2004). Innerhalb der Genetik und Evolutionstheorie kann durch Vergleich von Gensequenzen der Verwandtschaftsgrad durch biostatistische Verfahren bestimmt werden. Diese können mittels phylogenetischer Bäume und Netzwerke analysiert werden (Jungck and Weisstein, 2013). Auch die Verteilung von Restriktionsenzym Schnitten in einem Genom kann durch ein biostatistisches Verfahren berechnet werden.

## 2.2. Bioinformatik

In der Mitte des 20. Jahrhunderts wurden das erste Protein sequenziert (Sanger and Tuppy, 1951). Mit der Erforschung der DNA-Doppelhelix wurden in den folgenden Jahrzehnten immer mehr Proteine und nun auch DNA-Abschnitte sequenziert, welche zunächst innerhalb des *Protein Sequence Atlas* festgehalten wurden. Sequenzen wurden bereits am Anfang der 1970er Jahre mittels bioinformatischer Algorithmen verglichen. Ab dem Ende der 70er Jahren des 20. Jahrhunderts speicherte man immer größere werdenden Datenmengen in Datenbanksysteme (Schneider et al., 2010). So wurde im Jahre 1979 die Datenbank *Genbank* innerhalb der Los Alamos National Laboratory gegründet (Kanehisa et al., 1984). Es folgten viele weitere Datenbanken und viele DNA-, Protein- und sogar ganze Genomsequenzierungen. Auch die Algorithmen zum Sequenzvergleich wurden weiter entwickelt. Mittels des Alignment-Algorithmus (Needleman and Wunsch, 1970) oder des BLAST-Algorithmus (Altschul et al., 1990) können Gensequenzvergleiche durchgeführt und somit Genfunktionen oder die molekulare Phylogenie entschlüsselt werden.

Von den Anfängen, bei denen die Datenspeicherung Hauptaugenmerk der

Bioinformatik war, hat sich die Bioinformatik zu einer interdisziplinären weitgefächerten Wissenschaft weiter entwickelt. Die heutige Bioinformatik befasst sich mit der Sequenz-Genomanalyse, mit Phylogenie, Strukturbioinformatik, Analyse von Hochdurchsatzmethoden (“Omics”-Technologien) und anderen Auswertungen von Daten innerhalb biologischer Prozesse. Weitere Computer gestützten Verfahren innerhalb der Bioinformatik sind sehr vielfältig. Diese sind zum Beispiel Proteinstrukturvorhersagen, 3D-Visualisierung (Biasini et al., 2014), Genfindung, Medikamentendesign, Genexpressions-Prognosen, Protein-Protein-Interaktionen oder Clusteranalysen, dargestellt durch Venndiagrammen oder Dendrogramme.

Es gibt in der Biomathematik und der Bioinformatik viele Schnittstellen. Die Bioinformatik und die Biomathematik sind wichtige Teile der Systembiologie.

### 2.2.1. Sequenzdatenbanken

Bekannte DNA-Sequenzen zu bereits untersuchten Organismen werden innerhalb der Bioinformatik in Datenbanken gespeichert. Drei weltweit verbreitete Sequenzdatenbanken sind *ENA* (European Nucleotide Archive), *DD-BJ* (DNA Data Bank of Japan) und *Genbank*, siehe Tabelle 2.1. Innerhalb des *INSDC* (International Nucleotide Sequence Database Collaboration) werden seit 1992 die Inhalte von diesen drei primäre Sequenzdatenbanken täglich synchronisiert.

Im ersten Teil dieser Arbeit wird auf die *Genbank* des NCBI zurückgegriffen. Das NCBI (Cooper et al., 2010), *National Center for Biotechnology Information*, wurde am 4. November 1988 <sup>1</sup> gegründet. Verwaltet wird das NCBI vom LANL (Los Alamos National Laboratory).

---

<sup>1</sup> Gesponsert von Claude Pfeffer (★8. September 1900, †30. Mai 1989)

**Tabelle 2.1.: Primäre Sequenzdatenbanken**

Die innerhalb des INSDC zusammengefassten primären Sequenzdatenbanken aufgelistet.

Name	Institut	Ursprung
<b>Genbank</b>	National Center for Biotechnology Information (NCBI)	USA
<b>European Nucleotide Archive (ENA)</b>	European Bioinformatics Institute (EBI) und European Molecular Biology Laboratory (EMBL)	Europa
<b>DNA Data Bank of Japan (DDBJ)</b>	National Institute of Genetics (NIG)	Japan

Im Jahre 1979 wurde an dem LANL eine Sequenzdatenbank von Walter Goad (★1925, †2000) etabliert, welche drei Jahre später zur *Genbank* wurde (Kanehisa et al., 1984).

## 2.3. “Omics”-Technologien

Die “Omics”-Technologien (aus dem Englischen *Omics technology*) beinhalten die Charakterisierung von unterschiedliche biologischen Molekülen. Innerhalb der “Omics”-Technologien werden Strukturen, Funktionen und Dynamiken von den jeweiligen Molekülen untersucht (Attur et al., 2002). Hochdurchsatzmethoden sind bei den “Omics”-Technologien weit verbreitet. Hochdurchsatzmethoden, oder im Englischem High Throughput, sind biologische Methoden, in denen viele parallelisierte Experimente durchgeführt werden. Die Bioinformatik ist in der Lage, Daten von Hochdurchsatzmethoden auswerten.

Die “Omics”-Technologien werden je nach Molekülart unterteilt.

- **Genomik** (Gene). Die Genomik befasst sich mit dem Erbgut eines Organismus oder Virus, also der DNA (bzw. RNA). Die Genomik untersucht den Aufbau von Genomen und die Funktion und Wechselwirkungen von Genen.
- **Transkriptomik** (Transkripte / mRNA). Die Transkriptomik untersucht alle zu einer gewissen Zeit vorhandene Transkripte, also alle von DNA in RNA geschriebene Gene.
- **Proteomik** (Proteine, Proteom). Die Proteomik oder auch Translatomik beinhaltet die Erforschung des Proteoms zu einem Zeitpunkt, also aller Proteine, die innerhalb eines Organismus vorkommen.
- **Metabolomik** (Stoffwechselprodukte, Metabolite). Das Metabolom, welches alle charakteristischen Eigenschaften des Stoffwechsels eines Organismus oder Teilen eines Organismus beinhaltet, wird in der Metabolomik untersucht.
- **Fluxomik** (Flüsse von Metabolite). Das Fluxom beinhaltet die katalysierte Stoffumsetzungen und ihre Geschwindigkeit. Innerhalb der Fluxomik werden durch markierte Metabolite Flussraten und Kinetikparameter untersucht.
- **Glykomik** (Zucker). Die Glykomik bezeichnet die Untersuchung aller ZuckerMoleküle eines Organismus. Besonderer Fokus der Glykomik liegt bei Zuckermolekülen, die zusammen mit Proteinen interagieren und glykolysiert werden.
- **Lipidomik** (Lipide / Fette). Die Gesamtheit aller Lipide innerhalb einer Zelle und deren Aufgaben werden in der Lipidomik untersucht.



- **Interaktomik** (Interaktion von Molekülen). Das gesamte Netzwerk an Molekülen und deren Zusammenwirkung in einem Organismus oder einer Zelle ist Bestandteil der Interaktomik.
- **Phenomik** (Phenotypen). Die Phenomik erläutert alle Phenotypen eines Organismus. Besonderes Augenmerk liegt dabei auf der Veränderung des Phänotyps im Laufe eines Lebens.

Innerhalb der Biowissenschaften entstehen immer wieder neue Teilbereiche der “Omics”-Technologien.

## 2.4. Systembiologie

Die Aufgabe der Systembiologie ist es, ein abstrahiertes, aber möglichst holistisches Modell eines Organismus auf allen regulatorischen Ebenen zu entwickeln und dabei Vorhersagen zu ermöglichen (Ideker et al., 2001). Die Systembiologie soll ein umfassendes Verständnis der Interaktionen aller Bausteine eines biologischen Systems ermöglichen.

Historisch gesehen hat die Systembiologie ihren Ursprung in der Biomathematik, der Bioinformatik und den “Omics”-Technologien (Attur et al., 2002). Die Systembiologie greift heute als interdisziplinäre Wissenschaft sowohl auf viele unterschiedliche Bio-Wissenschaft, zum Beispiel Genetik, Biochemie, Molekularbiologie, Enzymologie, Evolutionstheorie oder Zellbiologie, als auch auf andere Wissenschaften, wie zum Beispiel Informationswissenschaften, mit Datenbanken-Systemen, Visualisierungs- oder auch Modellierungs-Werkzeuge, Mathematik wie Analysis, Statistik oder Modellierung, und anderen Lebenswissenschaften, wie zum Beispiel Chemie oder Physik, zurück.

Die Systembiologie untersucht komplexe Wechselwirkungen innerhalb eines biologischen Systems. Studien über Interaktionen von verschiedenen biologischen Komponenten können Aufschluss über Funktion und Verhalten einer Zelle bringen (Snoep et al., 2006). Durch verschiedene Hochdurchsatzverfahren wie Next-Generation-Sequencing oder Microarrays wurden in den letzten Jahrzehnten viele Details von biologischen Systemen entschlüsselt (Ideker et al., 2001). Einzelne Prozesse innerhalb von Zellen wurden charakterisiert, gespeichert, modelliert und zu einem gesamten biologischen System zusammengefasst. Dies geschieht zum Beispiel durch verschiedene statistische Auswertungsverfahren, mathematischen Modellierungen mittels dynamischer Systeme, computerunterstützten Simulationen oder Datenbanken.

Innerhalb der Systembiologie werden zwei Hauptansätze unterschieden (Brugeman et al., 2007):

- **Top-Down** beginnt mit einem Überblick über das System. Zunächst werden, ohne auf Details zu achten, über Hochdurchsatzmethoden viele experimentelle Daten gesammelt. Mittels dieser Daten soll auf die biologischen Mechanismen geschlossen werden. Da diese Modelle auf keiner mechanistischen Grundlage beruhen, sind sie meist phänomenologisch. Sie sind nützlich, um Einblicke in das System zu erhalten und somit Hypothesen zu formulieren.
- **Bottom-Up** ist der Ansatz bei dem von einem Modell ausgegangen wird, denen biologische Eigenschaften zugrunde liegen. Ein Netzwerk wird durch bekannte Eigenschaften, welche Information zu den kinetischen, physikalischen und chemischen Eigenschaften der involvierten Moleküle und deren Verhalten auf externe oder interne Signale sind, mechanistisch aufgebaut. Erst nachdem die Grundstruktur des Netzwerkes feststeht, wird dieses untersucht und modelliert, um damit Vorhersagen zu treffen.

## 2.5. Programmierung-Software

Innerhalb der Bioinformatik, der Biomathematik und der Systembiologie ist ein wichtiges Werkzeug, um Daten zu untersuchen, die Programmierung-Software. Diese Arbeit nutzt zwei unterschiedliche Programmiersprachen, Matlab und R. Beide Programmiersprachen sind unter den Top 30 der weitverbreitesten Programmiersprachen (nach TIOBE Index, Juli 2014<sup>2</sup>).

### 2.5.1. Matlab

Matlab (von *matrix laboratory*) ist eine von MathWorks entwickelte, numerische Programmiersprache. In den 1970er Jahren wurde es zunächst von Cleve Moler (★17. August 1939) in dem Fachbereich Informatik der Universität von New Mexico (USA) entwickelt. Zunächst wurde Matlab nur für Studenten entwickelt. Im Jahr 1984 gründete Cleve Moler mit Jack Little und Steve Bangert die Firma MathWorks und Matlab wurde zu einem kommerziellen Produkt.

Matlab dient der numerischen Lösung von mathematischen Problemen und unterstützt Matrizen-Rechnung, numerische Simulationen, Datenanalyse und -auswertung, Abbildung von Funktionen und Daten, Algorithmus-Implementierung, Schnittstellen zu anderen Programmiersprachen, z. B.: Fortran, Java, C und C++, und vieles mehr. Matlab ist eine weit verbreitete Programmiersprache, die heutzutage sowohl in der Industrie als auch an Hochschulen verwendet wird. Die Bereiche, in denen Matlab angewendet wird, sind sehr vielseitig. Dazu gehören zum Beispiel Wirtschaft, Ingenieurswesen oder Naturwissenschaften.

---

<sup>2</sup><http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>

## 2.5.2. R

R ist eine freie Programmiersprache für statistische Berechnungen, Datenbearbeitung und graphische Darstellung (Lafaye de Micheaux et al., 2011). R wurde von *The R Foundation for Statistical Computing* entwickelt. Die in R programmierten Algorithmen können auch in andere Programme oder Webseiten eingebunden werden.

R bietet viele zusätzliche Funktionen. Diese liegen in den verschiedensten Bereichen, nicht nur in der Statistik, sondern zum Beispiel auch zum Beispiel in der Numerik. Die Methoden von R beinhalten

- einfache Berechnung von Arrays und Matrizen,
- vielfältige Möglichkeiten zur Visualisierung von Daten und Ergebnissen,
- großen Sammlung von Paketen für die Datenanalyse (zum Beispiel *Bioconductor*),
- die Möglichkeit zur Erstellung von eigenen Paketen,
- verschiedene Interfaces zu anderen Programmiersprachen (*Python*, *C*, *C++*, *Fortran* oder *Perl*) und
- die leicht Erweiterbarkeit mittels Paketen.

Die statistische Auswertung von Daten ist essentieller Bestandteil der biologischen Forschung. In diesem Bereich bietet R eine große Anzahl an fertigen und getesteten Methoden, welche in Paketsammlungen, zum Beispiel in der *Bioconductor*-Paketsammlung, zu finden sind. Zu der Auswertung von Restriktionsreaktionen, Transposoninsertionen und deren Lokalisation wurde im ersten Teil dieser Arbeit R genutzt, um Programme zu entwickeln

und eine Statistik zu erstellen. Diese sind in Programm-Pakete zusammengefasst. Zudem können diese Pakete auch als Grundlage für weitere darauf aufbauende Programme genutzt werden.

In der Programmiersprache R wird mit unterschiedlichen Dateiformate gearbeitet, die biologische Daten beinhalten können.

- **Fasta-Format;** das Fasta-Format ist ein text-basiertes Format und kann entweder Nukleotid- oder Peptid-Sequenzen im IUPAC-Code beinhalten. Repräsentiert werden die Sequenzen durch einen *single-letter-code*. Jeder Buchstabe steht für ein Nukleotid. Dieses Format entstand aus dem Fasta-Software-Paket (Lipman and Pearson, 1985) und ist mittlerweile zum Quasi-Standard geworden. Durch den einfachen Aufbau des Fasta-Formates kann es leicht in andere Programme eingebunden werden.
- **Rd-Format;** Rd-Dateien sind R-Dokumentations-Dateien. Innerhalb der Programmiersprache R können Objekte abgespeichert werden. Diese werden in einer einfachen Struktur in Rd-Formaten abgespeichert. Dateien können innerhalb von R einfach eingebunden werden. Außerhalb von R kann die Rd-Datei durch Umformatierung in zum Beispiel Latex, HTML oder Text-Dateien genutzt werden.
- **gbk-Format;** ist ein flatfile-Format aus der Genbank (siehe Tabelle 2.1), welches für DNA-Sequenzen genutzt wird. Flatfiles sind einfach strukturierte Textdateien. Jede Zeile dieser Datei erhält mindestens einen Eintrag, wobei mehrere Einträge mit einem Trennzeichen getrennt sind. Dateien des gbk-Formats sind Textdateien und besitzen eine feste Breite der Absätze. Sie enthalten nicht nur die DNA-Sequenzen eines Bakteriums, sondern auch Information über Gene und Proteine, deren Form, Länge oder Erst-Entdeckung. Am Ende der Datei ist immer ein doppelter Schrägstrich.

Um Informationen zu einem Genom aus einer gbk-Datei in R einzulesen, wurde innerhalb der *Computational Microbiology*-Gruppe am Institut für Mikrobiologie der TU-Braunschweig ein R-Paket mit dem Namen *gparser* entwickelt. Das *gparser*-Programm ermöglicht direktes Zugreifen auf die DNA-Sequenz innerhalb einer Fasta-Datei.

## 2.6. Genregulation

Die Steuerung der Aktivität von Genen wird in der Biologie als Genregulation bezeichnet. In eukaryotischen Vielzellern ist die Regulation von Genen zum Beispiel notwendig, um eine Differenzierung einzelner Zellgruppen zu gewährleisten. Viele Einzeller, zum Beispiel Bakterien, leben unter sich stetig ändernden Umgebungsbedingungen. Gene müssen je nach Umgebung unterschiedlich stark exprimiert oder ganz aus-, bzw. angeschaltet werden. Dies kann durch die Genregulation beeinflusst werden.

Die Genregulation ist essentiell für Viren, Prokaryoten und Eukaryoten und zeigt eine große Bandbreite an Mechanismen. Zum Beispiel kann über die Modifikation der DNA, die Transkription, die Translation (Osterman et al., 2013), posttranslational, über den RNA-Transport oder den Abbau der mRNA regulatorisch eingegriffen werden. Innerhalb dieser Arbeit wird ausschließlich die transkriptionale Genregulation innerhalb Prokaryoten genauer betrachtet.

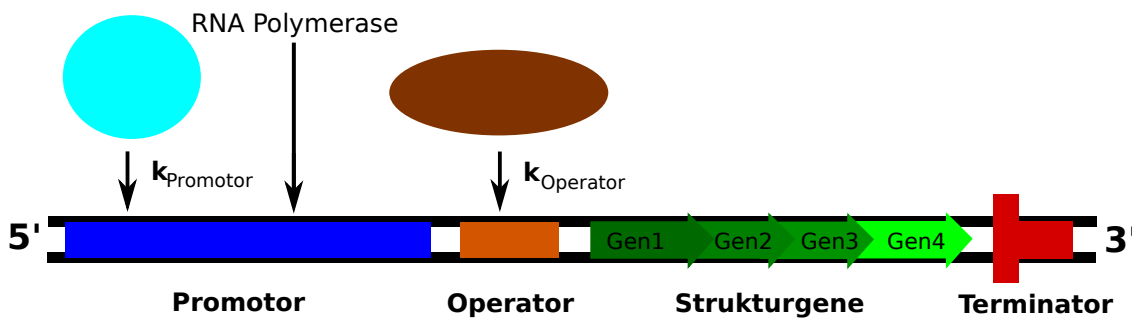
Transkriptionsregulation wird über spezielle DNA-Bindeproteine, den sogenannten Transkriptionsfaktoren, gesteuert. Diese Transkriptionsfaktoren binden an DNA-Bindestellen und regulieren so die Aktivität der RNA-Polymerase.

- **Repressoren** binden in der Nähe von oder direkt auf der Promotorregion (Operator) eines Gens. Repressoren verringern die Aktivität der

Genexpression oder verhindern diese ganz.

- **Aktivatoren** verbessern die Wechselwirkung zwischen der RNA-Polymerase und dem passenden Promotor. Es entsteht eine Förderung oder Aktivierung der Genexpression. Die meisten Regionen, an die die Aktivatoren binden, befinden sich in der Nähe der Promotorregion.
- **Allgemeine Transkriptionsfaktoren** (im Englischen: basal transcriptional factors) positionieren die RNA-Polymerase zu Beginn einer Protein-kodierenden Sequenz, wodurch die Expression der Gene starten kann.
- **Spezifische Faktoren** können die charakteristischen Eigenschaften der RNA-Polymerase verändern, wodurch an eine Reihe von Promotoren besser oder schlechter gebunden werden kann.

Innerhalb von Prokaryoten können Gene über Operons gesteuert werden. Operons bestehen aus einer DNA Sequenz, die in vier Abschnitte unterteilt wird (siehe Abbildung 2.1). Der Promotor ist (in 5'-3'-Richtung) der erste Teil des Operons und beinhaltet eine Bindestelle für die RNA-Polymerase und Bindestellen für Regulatoren, welche Aktivatoren oder Repressoren sein können. Diese kontrollieren die Aktivität der RNA Polymerase. Die Bindung dieser Regulatoren an den Promotor geschieht mit einer Rate, welche in Abbildung 2.1 mit  $k_{Promotor}$  bezeichnet wird. Aktivatoren binden meist im oder vor dem Promotor. Der Operator befindet sich hinter dem Promotor, an welchem Repressoren binden können. Dies geschieht mit einer Rate von  $k_{Operator}$ . Hinter dem Operator befinden sich die Strukturgene. Es ist immer mindestens ein Strukturgen innerhalb eines Operons, ein Operon kann aber auch mehr als ein Gen beinhalten. Der Terminator beendet die Transkription innerhalb eines Operons und liegt direkt hinter den Strukturgenen.



### Abbildung 2.1.: Aufbau eines prokaryotischen Operons

Ein Operon ist eine Transkriptionseinheit. In 5'-3'-Richtung wird der Promotor (blau), der Operator (orange), die Strukturgene (grün) und der Terminator (rot) dargestellt. An dem Promotor bindet die RNA-Polymerase und mit einer Bindungsrate von  $k_{Promotor}$  die Regulatoren. Innerhalb des Operators binden Repressoren mit einer Rate von  $k_{Operator}$ . Die Strukturgene befinden sich hinter dem Operator. Der Terminator beendet die Transkription des Operons.

Die Genregulation kann auch über mehrere Regulatoren, bzw. Transkriptionsfaktoren vermittelt werden, welche zusammen ein genregulatorisches Netzwerk bilden.

#### 2.6.1. Genregulatorische Netzwerke

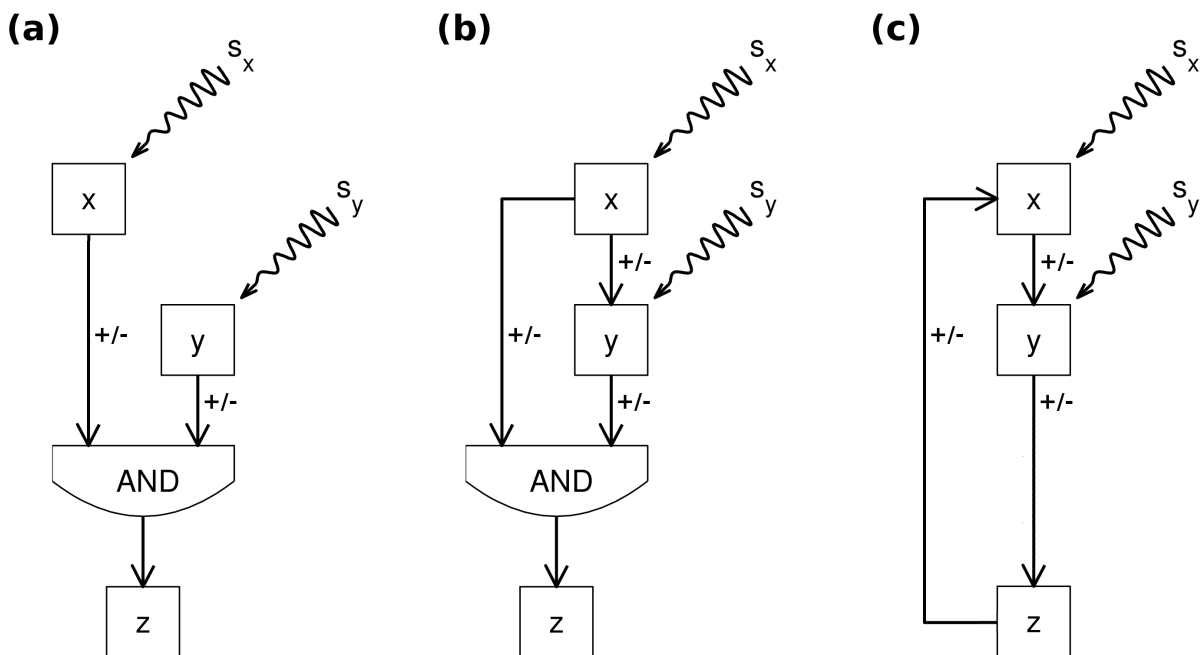
Um das Überleben unter immer wechselnden Bedingung zu sichern, ist es für Bakterien notwendig, mit regulatorischen Netzwerken die Expression von Genen zu steuern (Shen-Orr et al., 2002, Alon, 2007). Genregulatorische Netzwerke beinhalten Regulatoren, Sensoren und Zielproteine, bzw. Zielgene. Viele genregulatorischen Netzwerke werden zumeist durch Transkriptionsfaktoren gesteuert. Mittels Zwei-Komponenten-Systemen innerhalb von genregulatorischen Netzwerken werden Signale durch eine Sensorkinase detektiert. Danach werden die folgenden Gene mit einem Responsregulator gesteuert. Die Gene für die Sensorkinase und für den Responsregulator befinden sich auf einem Operon. Durch diese Zwei-Komponenten-System



können durch äußere Signale ganze Netzwerke gesteuert werden. Signale können reprimierend oder aktivierend wirken und nicht nur externe physikalische oder chemische Reize sein, sondern auch intern von der Zelle abgegeben werden. Sie resultieren in Aktivierung oder Reprimierung von Genen oder anderer Regulatoren.

Innerhalb von regulatorischen Netzwerken kann es zu Hoch- oder Runterregulierung oder sogar zur vollständigen An- und Abschaltung von Genen kommen. Ein Netzwerkmotiv ist ein Teil-Ausschnitt eines regulatorischen Netzwerkes. Es gibt verschiedene Aufbauten von regulatorischen Netzwerkmotiven. Neben der Autoregulation (Alon, 2007) ist einer der simpelsten Motive ein Signal, welches direkt die Expression eines Zielgens aktiviert oder reprimiert. Je mehr Regulatoren in einem Netzwerkmotiv involviert sind, desto komplizierter werden die Motive. Netzwerk motive mit drei Genen, mit den Genen  $x$ ,  $y$  und  $z$  (siehe Abbildung 2.2) sind sehr weit verbreitet. Drei Gen-Motiv-Netzwerke können in Untergruppen eingeteilt werden.

- Eine einfache *UND*-Schleife. Dabei aktivieren bzw. reprimieren zwei Regulatoren zusammen ein Zielgen (Alon, 2007). Das ist dargestellt in Abbildung 2.2 (a).
- Die Vorwärtskopplung (engl. feedforward Loop, FFL) besteht aus einer direkten und einer indirekten Regulation. Die Regulationen können aktivierend, wie auch reprimierend sein. Der erste Regulator steuert nicht nur die Produktion der Zielgene, sondern auch die Produktion des zweiten Regulators. Eine vollständig positive FFL, die sogenannte Typ-1 kohärente FFL, kommt in hunderten von Genregulationen in *Escherichia coli*, *Saccharomyces cerevisiae* oder anderen Organismen vor (Alon, 2007). Zum Beispiel befinden sich diese Netzwerke in der Regulation der Zuckernutzung in *E. coli* (Mangan et al., 2003, Mangan et al., 2006, Görke and Stülke, 2008), der Flagellen Expression



**Abbildung 2.2.: Drei-Gen-Netzwerkmotive in der Genregulation**

Drei verschiedene Regulationensmotive mit je 3 Genen ( $x$ ,  $y$  und  $z$ ) und zwei von außen kommenden Signalen ( $s_x$  und  $s_y$ ). Alle Regulationen innerhalb eines Motivs können Aktivierungen (+) oder Reprimierungen (-) sein, gekennzeichnet wurde dieses mit +/- . In Abbildung (a) befindet sich eine einfache UND-Regulation. Hierbei steuern zwei Regulatoren ( $x$ ,  $y$ ) direkt gemeinsam ein Zielgen ( $z$ ). Eine Feedforward-Schleife wird in Abbildung (b) dargestellt. Zwei Regulatoren können in der FFL auf einem indirektem und einem direktem Weg die Zielgene steuern. Der erste Regulator  $x$  beeinflusst dabei auch noch die Produktion des zweiten Regulators  $y$ . In Abbildung (c) wird die Feedback-Schleife dargestellt. Die Produktion des Zielgens ( $z$ ) hat eine Regulationsrückkopplung auf den ersten Regulator ( $x$ ).

in *Pseudomonas putida* (Kalir et al., 2005) oder der Regulation der Sporenbildung in *Bacillus subtilis* (Eichenberger et al., 2004). Auch von *Pseudomonas aeruginosa* wird diese aktivierende FFL zur Steuerung der Nitratatmung genutzt, was im zweiten Abschnitt dieser Arbeit genauer erläutert wird. In Abbildung 2.2 (b) wird eine FFL mit drei Genen dargestellt.

- Eine Rückkopplung (engl. feedback loop) kann ebenfalls sowohl eine

aktivierende als auch reprimierende Steuerung sein. Sowohl negative Feedback-Schleife (NFL, engl. negative feedback loop) als auch positive Feedback-Schleife (PFL, engl. positive feedback loop) kommen häufig in der Genregulation vor (Alon, 2007, Mitrophanov and Groisman, 2008). Bereits in vielen Studien wurden Funktionen von NFLs und PFLs beschrieben (Becskei et al., 2001, Hornung and Barkai, 2008, Stekel and Jenkins, 2008). Eine negative Feedback-Schleife (NFL) kann die gesamte Konzentration verringern (Oehler et al., 1990). Eine positive Feedback-Schleife kann ein bistabiles Verhalten verursachen (Hasty et al., 2002). Dargestellt wird die FBL in Abbildung 2.2 (c).

Regulatorische Netzwerke können mittels mathematischer Modelle genauer untersucht werden. Hierzu werden zumeist gewöhnliche Differentialgleichungen (engl. ordinary differential equations - ODEs) genutzt (siehe Kapitel 2.7).

## 2.7. Modellierung von Genregulationen über Differentialgleichungen

Mittels des Wissens über den Aufbau der genregulatorischen Netzwerke und Reaktionsraten können Genregulationen durch Differentialgleichungssystemen modelliert werden. Reaktionskinetiken beschreiben die Geschwindigkeiten von Reaktionen und die Reaktionsraten von chemischen Prozessen. Substrate werden zu Produkten umgesetzt. Substrate und Produkte werden als Reaktionsgrößen bezeichnet.

Die Reaktionsgeschwindigkeit einer Reaktionskinetik gibt an, wie viele Teilchen pro Zeit umgesetzt werden. Die Geschwindigkeit kann sowohl von äußeren Bedingungen abhängen, wie zum Beispiel Sauerstoffkonzentration,

Temperatur, Salzgehalt, pH-Wert, etc., als auch von der Stoffmenge selbst. Je mehr Teilchen vorliegen, desto mehr Teilchen können auch umgesetzt werden. In Zellen werden chemische Reaktionen meist durch Enzyme katalysiert.

**Definition 1.** Eine Gleichung

$$y = f(x, y, y', y'', \dots, y^n),$$

welche die Funktion  $y = y(x)$  und dessen Ableitung bis zur  $n$ -ten Ordnung enthält, heißt **gewöhnliche Differentialgleichung  $n$ -ter Ordnung**.

Die Funktion  $y = y(x)$  ist dann die Lösung der gewöhnlichen Differentialgleichung, wenn sie die gewöhnliche Differentialgleichung mit all ihren Ableitungen erfüllt. Hierbei ist zwischen der allgemeinen Lösung und der expliziten Lösung zu unterscheiden. Die explizite Lösung lässt sich dabei durch Anfangs- und Randbedingungen analytisch oder numerisch berechnen.

Innerhalb von gewöhnlichen Differentialgleichungen treten Ableitungen der gesuchten Funktion nur nach genau einer Variable auf. Hängt die gesuchte Funktion von mehreren Variablen ab und treten Ableitungen nach verschiedenen Variablen auf, so spricht man von partiellen Differentialgleichungen. Innerhalb dieser Arbeit werden nur gewöhnliche Differentialgleichungen genutzt.

Die Konzentration jedes in die Reaktion involvierten Stoffes wird angegeben durch  $x_i$  mit  $i \in \{1, \dots, n\}$  und  $n$  der Anzahl an allen Komponenten. Die Funktion  $f_i$  bildet den Zustand des Systems der Stoffe  $\{x_1, \dots, x_n\}$  auf die Veränderung von  $x_i$  ab. Die Veränderung

$$\dot{x}_i = f_i(x_1, \dots, x_n), \quad \forall i \in \{1, 2, \dots, n\}$$

wird für jeden Stoff  $x_i$  angeben, wobei  $f_i$  durch die Reaktionsraten ermittelt werden kann. Diese Raten können auch von anderen Stoffen abhängig sein. Auch regulatorische Netzwerke lassen sich über Reaktionskinetiken und Differentialgleichungsmodelle untersuchen.

Das mathematische Modell eines Regulations-Netzwerkes kann zum Beispiel mit Hilfe von Kinetiken oder der Hill-Funktion erstellt werden.

Die Michaelis-Menten-Kinetik (nach Maud Menten<sup>3</sup> und Leonor Michaelis<sup>4</sup>) beschreibt auf einfache Art die Reaktion von Enzymen durch die Reaktionsrate

$$v = \dot{P} = V_{\max} \frac{S}{K_m + S} \quad (2.1)$$

in Abhängigkeit des Substrates  $S$ , der maximalen Reaktionsrate  $V_{\max}$  und der Michaelis-Menten-Konstante  $K_m$  (Deichmann et al., 2014). Michaelis-Menten-Konstante ist die Konzentration des Substrats  $S$  bei der die Produktionsrate halbmaximal ist.

Ist die Bindung eines neuen Liganden, bzw. Substrat an ein zentrales Molekül abhängig davon, wie viele bereits gebunden sind, so spricht man von dem Phänomen der Kooperativität. Der Hill-Koeffizient  $h$ , nach Archibald Vivian Hill<sup>5</sup>, gibt die Stärke der Kooperativitäten an. Zumeist wird der Einfachheit halber ein natürlicher Hill-Koeffizient,  $h \in \mathbb{N}$ , genutzt. Ist  $h = 1$  liegt keine Kooperativität vor, die Stärke der Bindung der Liganden ist unabhängig voneinander. Der Hill-Koeffizient steigt mit größer werdener Kooperativität. Die Hill-Funktion zeigt einen ähnlichen Aufbau wie die Michaelis-

---

<sup>3</sup>Menten - ★16. Januar 1875, †8. Oktober 1949

<sup>4</sup>Michaelis - ★20. März 1879, †26. Juli 1960

<sup>5</sup>Hill - ★26. September 1886, †3. Juni 1977

Menten-Kinetik, hierbei ist jedoch

$$\theta = \frac{S^h}{K_b^h + S^h} \quad (2.2)$$

der Anteil an Bindestellen mit gebundenem Liganden, bzw. Substrat. Die Bindekonstante  $K_b$  ist die Konzentration bei halber maximaler Reaktionsrate und  $S$  die Menge an freiem Substrat.

## 2.8. Ziel der Arbeit

In dieser Arbeit sollen experimentelle Beobachtungen durch biomathematischen Modellierungen und bioinformatischen Analysen verifiziert und wenn möglich Aussagen verfeinert werden. Diese sollen mittels erneuter Experimente bestätigt werden. Innerhalb von drei Themengebiete werden Modell und Analysen aufbauend auf mikrobiologischen Beobachtungen aufgestellt. Ein Teil befindet sich innerhalb der Bioinformatik. Zwei Themengebiete befinden sich in der mathematischen Modellierung von Genregulationen.

Innerhalb des ersten Teils wird in Eigenarbeit ein Algorithmus entwickelt und implementiert, welcher Transposoninsertionen in bakteriellen Genomen mit geringem Aufwand lokalisieren kann. Der Algorithmus, *InFiRe* genannt, basiert auf der eindeutigen Verteilung von Restriktionsenzymsschnitten. Mittels *InFiRe* sollen innerhalb der Transposoninsertionssuche langwierige Sequenzierungsarbeiten entbehrlich gemacht werden. Der Transposoninsertion können durch Beobachtungen des Southern-Blot-Hybridisierung nach Restriktionsverdau mehrere mögliche Lokalisationen zugeordnet werden. Durch den *InFiRe*-Algorithmus soll die Anzahl der möglichen Positionen reduziert, daher die Beobachtung verfeinert werden. Sequenzierung

kann das Ergebnis bestätigen. Die Lokalisierung von Transposoninsertionen kann die Charakterisierung von Genen unterstützen. Der Algorithmus kann als Grundlage für ein Hochdurchsatzverfahren innerhalb der Genomik dienen. *InFiRe* soll eine kostengünstige und zeitsparende Alternative zu bereits vorhanden Verfahren der Transposoninsertionssuche darstellen.

Im zweiten und dritten Teil werden Beobachtungen der Genregulationen innerhalb von Bakterien modelliert und analysiert. Aufbauend auf experimentellen Daten werden zwei Regulationssysteme mittels mathematischer Modellen untersucht.

In Teil 2 soll ein Regulationsmotiv innerhalb von *Pseudomonas aeruginosa* beobachtet werden. *P. aeruginosa* kann sowohl mit, wie auch ohne Sauerstoff überleben. Mittels eines Regulationsnetzwerkes passt sich das Bakterium der jeweiligen Umwelt an. In diesem Regulationssystem kann man ein weitverbreitetes Regulationsmotiv beobachten, die sogenannte Feedforward-Schleife (FFL). Diese liegt hier jedoch mit einer Spezialisierung, nämlich einer dauerhaften basalen Expression eines Regulators, vor. Diese dauerhafte Expression des Regulator-Gens kostet das Bakterium Energie, bereitet es jedoch besser auf Umweltwechsel vor. Durch eine Modellierung der Regulation mittels eines Differentialgleichungssystems wird eine Nutzenbilanz-Optimierung durchgeführt und bestimmt, welchen Vorteil diese basale Expression für das Bakterium hat und unter welchen Umweltbedingungen sie vorteilhaft ist.

Der globale Virulenzregulator von *Yersinia pseudotuberculosis*, RovA, wird ebenfalls mittels eines Regulationsmotives gesteuert. Dieses Motiv beinhaltet eine reprimierende und eine aktivierende Feedback-Schleife. Es wurde beobachtet, dass die Produktion von RovA eine temperaturabhängige Bistabilität aufweist. Innerhalb des dritten Teils dieser Arbeit soll die Regulation

der Infektion von *Y. pseudotuberculosis* mittels Differentialgleichung analysiert werden, wobei die Werte der Parameter direkt aus experimentellen Daten bezogen werden sollen. Die zentrale Fragestellung dieses Teils der Arbeit ist die, wie sich Veränderungen des Modells auf die Bistabilität auswirken und wie robust das System ist.



**Teil I.**

***InFiRe* - Identifikation von  
Transposon-Insertions-  
stellen in bakteriellen  
Genomen**



### 3. Einleitung

Eine etablierte Methode in der funktionellen Genomik ist die Transposonmutagenese. Transposons sind mobile genetische Elemente, die zufällig in einem Zielgenom inserieren und dort Gene inaktivieren können. Dabei hinterlassen sie eine eindeutige Sequenz-Signatur. Der limitierende Schritt des Verfahrens ist die Lokalisierung der Transposonsignatur innerhalb eines Genoms. Bereits entwickelte Methoden sind meist sehr zeit- und kostenintensiv. Der im Rahmen dieser Arbeit entwickelte Algorithmus und die dazugehörige Software *InFiRe*, Insertion Finder via Restriction digestion, ist eine Computerunterstützte Alternative zu den bisherigen Methoden für das Orten von Transposoninsertion bei sequenzierten bakteriellen Genomen.

Die in Eigenarbeit geschriebene *InFiRe*-Software basiert auf der Analyse der Verteilung der Fragmentlängen nach einem Restriktionsverdau der bakteriellen DNA mittels verschiedener Restriktionsenzyme (Shevchuk et al., 2012). Über eine Southern-Blot-Hybridisierung mit einer Transposon spezifischen Sonde kann die Länge des Fragments mit Transposoninsertion ermittelt werden. Der beobachteten Größe des Fragments können mehrere Positionen innerhalb des Genoms zugeordnet werden. Durch wiederholte Restriktionsverdau und mittels *InFiRe* kann die Anzahl an möglichen Fragmentpositionen reduziert werden.

Mittels einer *mini-Tn10* Mutationsbibliothek von *Legionella pneumophila* Corby (Shevchuk et al., 2009) wurde die Funktionalität von *InFiRe* verifi-

ziert. Das Programm ist sowohl auf einer Webseite <sup>1</sup> ausführbar als auch als R-Paket <sup>2</sup> verfügbar.

## 3.1. Transposon

Transposons wurden zum ersten mal 1948 von Barbara McClintock (★16. Juni 1902, †2. September 1992) während ihrer Forschung am Mais entdeckt (McClintock, 1950, McClintock, 1953).

Transposons sind mobile genetische Elemente, die aus DNA oder RNA bestehen. Sie können zufällig in ein Genom von Viren, Prokaryoten oder Eukaryoten inserieren (Hayes, 2003, Largaespada, 2009, Vilen et al., 2003, Yergeau and Mead, 2009). Transposons haben auch die Möglichkeit, innerhalb eines Genoms ihre Position zu verändern und nicht nur auf Chromosomen, sondern auch auf Plasmide zu springen (Clewell, 1981). Transposoninsertionen können Gene inaktivieren und haben somit auch evolutionär und innerhalb der Forschung eine große Bedeutung (Clewell, 1981).

### 3.1.1. Transposon-Klassen

Transposons werden in zwei Klassen unterteilt (Wicker et al., 2007). Transposons der Klasse 1 werden kopiert und die Kopie wird andernorts inseriert, Transposons der Klasse 2 werden ausgeschnitten und dann an neuer Stelle eingefügt.

---

<sup>1</sup>[www.InFiRe.tu-bs.de/analyze](http://www.InFiRe.tu-bs.de/analyze)

<sup>2</sup>[www.InFiRe.tu-bs.de/download](http://www.InFiRe.tu-bs.de/download)

## Klasse 1

Die Transposons der Klasse 1 werden auch Retrotransposons genannt (Sanmiguel and Bennetzen, 1998). Die DNA der Retrotransposons wird zunächst zu RNA und dann in zurück zu cDNA durch eine im Transposon meist enthaltene reverse Transkriptase transkribiert. Die cDNA, welche der DNA des Transposons entspricht, wird dann in das Genom an einer neuen Position inseriert.

## Klasse 2

Transposons der Klasse 2 werden als DNA-Transposon bezeichnet. Die DNA der Transposons wird mittels des Transposase-Enzyms aus dem Genom geschnitten und an einem neuen Ort eingefügt. Eine DNA-Polymerase schließt die Lücken und eine DNA-Ligase fügt das Transposon an einem neuen Ort ein.

### 3.1.2. Transposon-Mutagenese

Die Möglichkeit, mit DNA-Transposons Gene ausschalten zu können, wird Transposon-Mutagenese genannt. Sie ist eine fundamentale Möglichkeit zur Erforschung von Genfunktionen. Es gibt viele Methoden und neue Techniken, welche die traditionellen Transposon-Mutagenese erweitern. Diese sind zum Beispiel

- Signatur Ziel-Mutagenese (Hensel et al., 1995),
- Transposon vermittelte differentiale Hybridisierung (Chaudhuri et al., 2009),
- Genetisches *Footprinting* (Smith et al., 1995) oder

- Gene-Expression Analyse (Judson and Mekalanos, 2000, Opperman et al., 2003).

Der schwierigste, aber zugleich essentielle Teil dieser Methoden ist die am Anfang stehende Lokalisierung der Transposoninsertionsstellen. Bisherige Methoden zur Lokalisierung der Transposoninsertionen sind zum Beispiel Vectorette PCR (Arnold and Hodgson, 1991), inverse PCR (Ochman et al., 1988) oder Single-primer PCR (Ma et al., 2011). Hierbei ist die Single-primer PCR eine Methode, die ein schnelles Identifizieren der Insertionsstelle erlaubt. Grundvoraussetzung für diese Methode sind jedoch optimale Amplifikationsbedingungen.

## 3.2. Restriktionsenzyme

Zum ersten Mal kam der Begriff des Restriktionsenzyms in den 1950er Jahren im Zusammenhang mit einer Studie über den  $\lambda$ -Phagen auf (Luria and Human, 1952). Restriktionsenzyme sind Enzyme, welche in DNA schneiden können (Bickle, 2004). Dabei hat jedes Restriktionsenzym vom Typ II (siehe Tabelle 3.1) eine eindeutige Basenreihenfolge, bei der das Enzym schneidet, die sogenannte Erkennungssequenz. In der Natur treten Restriktionsenzyme zumeist bei Bakterien oder Archaeen auf, die diese zur Phagenabwehr nutzen. In der Forschung werden Restriktionsenzyme genutzt, um DNA an bestimmten Stellen zu schneiden, zum Beispiel, um Vektoren zu linearisieren oder um Teile des Genoms zu entfernen.

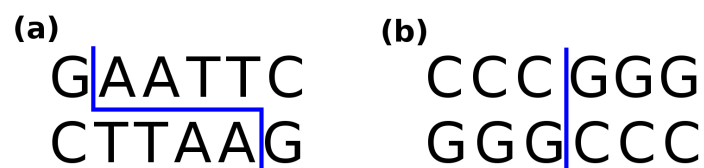
Restriktionsenzyme teilen sich in 4 verschiedene Klassen auf (Bickle and Krüger, 1993). Diese werden in Tabelle 3.1 gezeigt.

**Tabelle 3.1.: Typen von Restriktionsenzyme**

Es gibt insgesamt vier unterschiedliche Typen an Restriktionsenzyme.

Typ	schneidet
<b>Typ I</b>	zufällig, weit weg von Erkennungssequenz, benötigt ATP
<b>Typ II</b>	in der Nähe oder innerhalb Erkennungssequenz
<b>Typ III</b>	~ 25 Bp entfernt von Erkennungssequenz, benötigt ATP
<b>Typ IV</b>	methylierte, hydroxymethylierte, etc. DNA

Zumeist werden innerhalb der Forschung Restriktionsenzyme des Typs II genutzt, da die Erkennungssequenzen exakt definiert sind (Pingoud et al., 2005). Innerhalb der Typ II Restriktionsenzyme wird durch die Schnittart unterscheiden (siehe Abbildung 3.1). Es gibt zwei verschiedene Arten. Das Restriktionsenzym vom Typ II kann so genannte *klebrige* Enden, *sticky ends* genannt, oder *glatte* Enden, *blunt ends* genannt, produzieren (Goodsell, 2002). Restriktionsenzyme, die *glatte* Enden produzieren, schneiden die beiden DNA-Stränge an exakt der gleichen Stelle. So gibt es keine Überhänge der einzelnen DNA-Stränge. Bei *klebrigen* Enden werden die DNA-Stränge mit vier Basen versetzt geschnitten. Beide Seiten des Schnittes haben einen Strang mit vier Basen Überhang. Die DNA-Enden mit *klebrigen* Enden können schneller ligieren, da die überlappenden DNA-Stränge miteinander hybridisieren können.

**Abbildung 3.1.: Schnittarten von Restriktionsenzymen**

Gezeigt werden zwei verschiedene Möglichkeiten von Restriktionsenzymen in DNA zu schneiden. Der blaue Strich deutet an, wo geschnitten wird. In (a) werden so genannte '*sticky ends*', also klebrige Enden, von dem Restriktionsenzym produziert. In (b) befindet sich ein Beispiel für '*blunt ends*', also glatte Enden.

Häufig besteht die Erkennungssequenz von Restriktionsenzymen des Typs II aus Palindromen mit vier, sechs oder acht Basen (Pingoud et al., 2005). Je länger die Erkennungssequenz ist, desto seltener schneiden die Restriktionsenzyme und hat somit Einfluss auf die Fragmentlängenverteilung (siehe Kapitel 4.1).

### 3.3. Southern Blot

Auf der Methode der Agarose-Gelelektrophorese aufbauend entwickelte im Jahre 1975 Edwin Southern (★7.Juni 1938) eine Methode zur Untersuchung von DNA-Fragmenten. Diese Methode wird Southern Blot genannt (Southern, 1975, Southern, 2006).

Bei einem Southern Blot wird zunächst die zu untersuchende DNA mittels Restriktionsenzymen (siehe Kapitel 3.2) in Fragmente geschnitten und in einer Gelelektrophorese der Länge nach getrennt und durch Alkalien denaturiert. Diese nun als Einzelstränge vorliegende DNA wird zumeist auf Nylonmembranen übertragen und dort fixiert (Southern, 2006). Hybridisiert man den zu suchende Sequenz auf der DNA mit einer markierten Sonde, kann der Abschnitt detektiert werden.



## 4. Aufbau und Methoden des *InFiRe*-Algorithmus

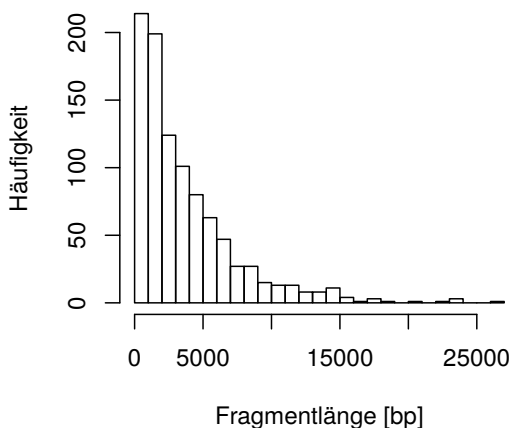
Mit dem Ziel, die Effizienz von Transposon-Mutagenese-Experimenten in bakteriellen Genomen zu erhöhen, wurde der in Eigenarbeit programmierte Algorithmus *InFiRe* entwickelt. Der Algorithmus ermittelt aus der individuellen Verteilung von Restriktionsenzym-Schnitten über eine Southern-Blot-Hybridisierung (Kapitel 3.3) die Transposoninsertions-Stelle. Der Algorithmus ist in zwei in R programmierten Paketen aufgeteilt, welche *Digestion* und *InFiRe* genannt werden. Die Anzahl der benötigten Restriktionsreaktionen kann mit zwei eigens etablierten Statistiken ermittelt werden.

Eine Grundvoraussetzung für den *InFiRe*-Algorithmus und der dazugehörigen Software ist die Fragmentlängenverteilung nach einer Restriktionsreaktion. Die Verteilung der Fragmentlängen kann mit der Genomsequenz eines Bakteriums und die Erkennungssequenz der Restriktionsenzyme ermittelt werden. Bakterielle Genomsequenzen sind unter anderem auf der Webseite des NCBI (siehe Kapitel 2.2.1) verfügbar. Damit lässt sich ein Algorithmus erstellen, bei dem man die Stelle der Transposoninsertion ermitteln kann.

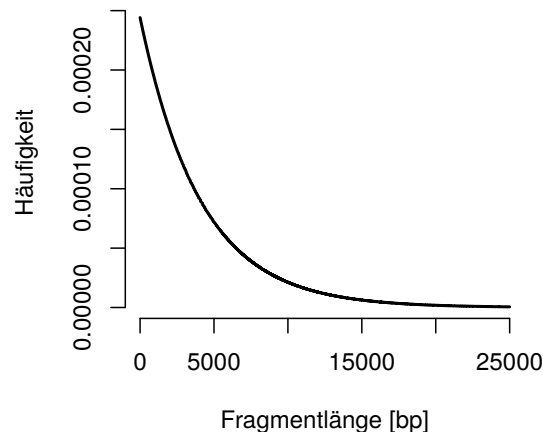
## 4.1. Fragmentlängenverteilung

Nach einem Restriktionsverdau einer bakteriellen DNA entstehen durchschnittlich  $S$  kleineren DNA-Fragmente, dessen Fragmentlänge mit  $x$  und die durchschnittliche Fragmentlänge mit  $m$  bezeichnet wird. Innerhalb der bakteriellen DNA kommen vier verschiedene Nukleinbasen ( $p = 4$ ) vor. Diese werden mit dem IUPAC-Code abgekürzt, Adenin ( $A$ ), Guanin ( $G$ ), Cytosin ( $C$ ) und Thymin ( $T$ ) (Nomenclature, 1970). Es wird angenommen, dass alle Nukleinbasen gleichhäufig und zufällig innerhalb eines Genoms vorkommen. Die Fragmentlängen haben eine näherungsweise exponentielle Verteilung (Grenet and Robelin, 1988) (siehe Abbildung 4.1 (a) + (b)). Nur durch die individuelle Verteilung von Fragmentlängen  $x$  kann der *InFiRe*-Algorithmus arbeiten.

(a)



(b)



**Abbildung 4.1.: Exponentialverteilung der Fragmentlängen nach einem Restriktionsverdau und theoretische Exponential Verteilung**

(a) Häufigkeitsverteilung der Fragmentlängen nach einer Restriktionsreaktion. Hier wurde *Legionella pneumophila* Corby mit dem Restriktionsenzym *EcoRI* geschnitten.

(b) Theoretische exponentielle Dichtefunktion mit der Funktion  $f(x) = \frac{1}{p^\ell} e^{-x/p^\ell} = \frac{1}{4096} e^{-x/4096}$

Die durchschnittliche Fragmentlänge  $m$  kann mittels der Länge der Erkennungssequenz  $\ell$  des Restriktionsenzym und der Anzahl an unterschiedlichen Basen  $p$  berechnet werden.

**Satz 1.** *Die durchschnittliche Fragmentlänge  $m$  wird mit*

$$m = p^\ell \quad (4.1)$$

*berechnet.*

*Mit der Länge der Genomsequenz eines Bakteriums  $n$  und der durchschnittlichen Fragmentlänge  $m$  kann die durchschnittliche Anzahl an Schnitten eines Restriktionsenzym*

$$S = \frac{n}{d} \quad (4.2)$$

*berechnet werden.*

Je größer eine Fragmentlänge  $x$  ist, desto kleiner ist die Anzahl an Fragmenten mit dieser Länge. Der Zusammenhang der Fragmentlänge und der Anzahl an Fragmenten ist durch eine exponentielle Verteilung gegeben.

**Satz 2.** *Die theoretische Dichtefunktion  $P_1$  von Fragmentlängen hängt von der durchschnittlichen Fragmentlänge  $x$  ab. Sie wird berechnet mit*

$$P_1(x) = \frac{1}{p^\ell} e^{\frac{-x}{p^\ell}}, \quad (4.3)$$

*unter der Voraussetzung zufällig verteilter Basenpaare.*

**Beweis 1.** Die theoretische Dichtefunktion der Fragmentlängen ist die Dichte einer Exponentialverteilung mit einer Ereignisrate von  $\frac{1}{\lambda} = m = p^\ell$ , also  $\lambda = \frac{1}{p^\ell}$ .

□

**Beispiel 1.** Eine DNA-Sequenz besteht aus einem Code mit vier verschiedenen Nukleinsbasen,  $p = 4$ . Es wird angenommen, dass die vier Basen gleich häufig im Genom vorkommen. Das Restriktionsenzym *EcoRI* hat eine Basenerkennungssequenz mit einer Länge von sechs Nukleinsbasen ( $\ell = 6$ ). Die durchschnittliche Fragmentlänge ( $m$ ) ist somit bei diesem Beispiel

$$m = p^\ell = 4^6 = 4096.$$

Die theoretische Dichtefunktion

$$P_1(x) = \frac{1}{4096} e^{\frac{-x}{4096}}$$

der Fragmentlängen wird in Abbildung *refpic:theoVerteilungInFiReI* (b) dargestellt. Das bakterielle Genom von *Legionella pneumophila* Corby hat  $n = 3576470$  Basenpaare. Die durchschnittliche Anzahl an Schnitten von *Legionella pneumophila* Corby mit *EcoRI* ist

$$S = \frac{n}{m} = \frac{3576470}{4^6} \approx 873.$$

## 4.2. Fragmentsuche in bakterieller DNA nach einem Restriktionsverdau

Das Genom kann als Basensequenz-Folge betrachtet werden. Die Basensequenz-Folge wird mit

$$z = (z_h)$$

bezeichnet und ist eine Folge der vier Basen,  $z_h \in \{A, T, G, C\}$  mit  $h \in \{1, \dots, n\}$  und  $n$  der Länge der Sequenz. Die Restriktionsenzyme des Typs II schneiden bei einer bestimmten Basenfolge in der DNA (siehe Kapitel 3.2). Diese

Basenfolge ist die Erkennungssequenz

$$z^R = (z_h)^R$$

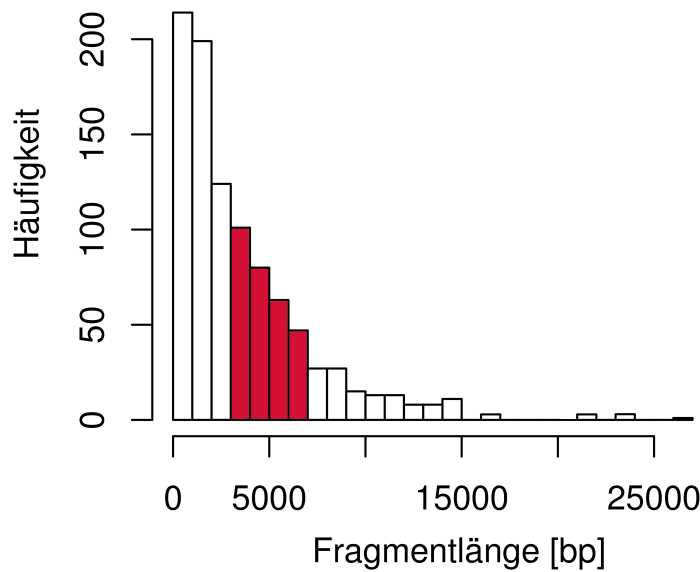
mit  $h \in \{1, \dots, \ell\}$ ,  $\ell$  der Länge der Erkennungssequenz und  $R$  dem passenden Restriktionsenzym.

Ein Genom unterteilt sich nach einem Restriktionsverdau in  $S$  neue kleinere Fragmente, welche mit  $z^s$  bezeichnet werden, wobei  $s \in \{1, \dots, S\}$ . Jedes Fragment  $z^s$  kann innerhalb des Genoms  $z = (z_h)$  lokalisiert werden und hat eine dazugehörige Fragmentlänge  $x \in X$ , mit  $X$  der Menge aller möglichen Fragmentlängen und einer durchschnittlichen Fragmentlänge von  $m$  (siehe Kapitel 4.1). Es ist eine Zuordnung von dem Fragment zu dessen Länge  $L(z^s) = x$  möglich. Die Umkehrfunktion

$$L^{-1} : X \rightarrow M, \quad \text{mit } z^s \in M \tag{4.4}$$

kann pro Fragmentlänge  $x$  mehrere Fragmente  $z^s$  zuordnen, mit  $s \in \{1, \dots, S\}$ . Die Fragmente  $z^s$  innerhalb der Menge  $M$  sind ebenfalls Folgen von Basenpaaren mit  $z^s = (z_h)^s$  mit  $z_h \in \{A, T, G, C\}$  mit  $h \in \{1, \dots, x\}$  und  $s$  dem dazugehörigen Fragment.

Ist die exakte Genomsequenz eines Bakteriums bekannt, so können die Schnittstellen einer Restriktionsendonuklease exakt bestimmt und somit auch die exakte Längenverteilung ermittelt werden. Durch  $L^{-1}$  kann eine Menge an Fragmenten  $z^s \in M$  mit bekannter Stelle im Genom zu der Fragmentlänge  $x$  zugeordnet werden (siehe Gleichung (4.4)).



**Abbildung 4.2.: Häufigkeitsverteilung der Fragmentlängen nach einem Restriktionsverdau mit Fragment-Suche.**

Hier wird ein Verdau dargestellt von *Legionella pneumophila* Corby mit dem Restriktionsenzym *EcoRI*. Wird ein Fragment mit einer Länge von 5000 bp gesucht, so kommen alle Fragmente auf dem Intervall von [4000, 6000] bp in Frage. Diese Fragmente werden hier in rot markiert. Jedem dieser Fragmente kann ein Ort im Genom zugeordnet werden. Alle rot markierten Fragmente können der Menge  $\bar{M}$  zugeordnet werden.

Die Fragmentlänge  $x$  ist nur ungefähr über eine Southern-Blot-Hybridisierung bestimmbar. Es tritt eine Ableseungenauigkeit  $a$  auf. Allen Fragmenten, die eine Fragmentlänge auf dem Intervall  $[x - a \cdot x, x + a \cdot x]$  haben, können das gesuchte Fragment sein und werden somit der Menge der möglichen Fragmente  $\bar{M}$  zugeordnet. Die Menge aller möglichen Fragmente ist somit explizit

$$\bar{M} = \int_{x-a \cdot x}^{x+a \cdot x} L^{-1}(t) dt.$$

In Abbildung 4.2 wird ein Verdau von *Legionella pneumophila* Corby mit dem Restriktionsenzym *EcoRI* gezeigt. Wird ein Fragment gesucht, welches die Länge von  $x = 5000$  bp haben soll und eine Ableseungenauigkeit von

$a = 0.2$  vorliegt, kommen alle Fragmente auf dem Intervall  $[x - a \cdot x, x + a \cdot x] = [5000 - 0.2 \cdot 5000, 5000 + 0.2 \cdot 5000] = [4000, 6000]$  bp in Frage. Dieses Intervall ist in der Abbildung 4.2 rot markiert. Man kann diesem Intervall der Fragmentlänge nun direkt Fragmente innerhalb des Genoms und dessen Position zuordnen.

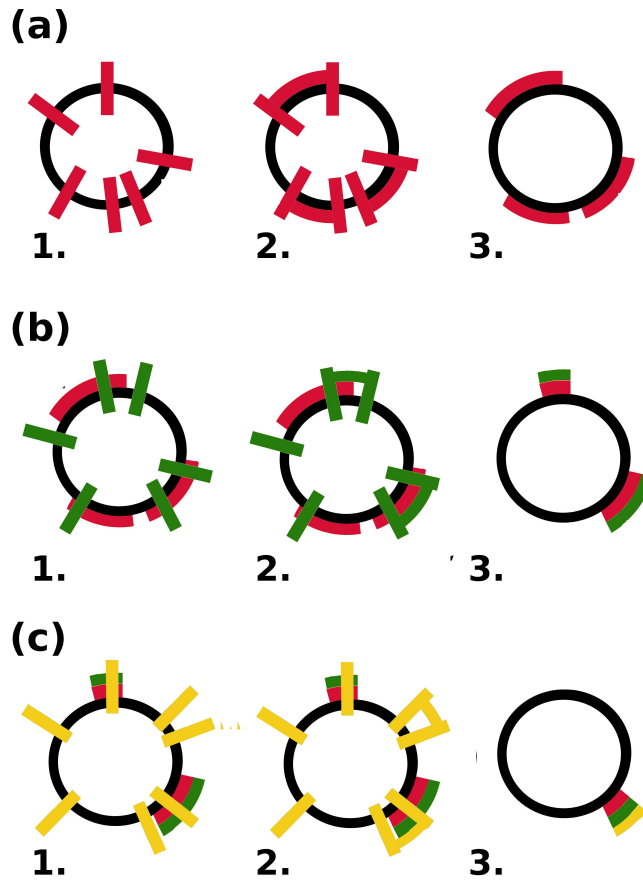
### 4.3. InFiRe-Algorithmus

Wird eine Transposoninsertion innerhalb des Genoms gesucht, kann nach einem Restriktionsverdau die Fragmentlänge  $x$  des Fragments auf dem sich die Transposoninsertion befindet mittels Southern-Blot-Hybridisierung bestimmt werden. Allen Fragmenten mit der abgelesene Fragmentlänge  $x$  mit einer Ableseungenauigkeit  $a$  wird Positionen innerhalb des bakteriellen Genoms zugeordnet. Alle möglichen Positionen befinden sich in der Menge  $\bar{M}$  (siehe Kapitel 4.2).

Gegeben sind Restriktionsverdaue, mit  $j \in \{1, \dots, i\}$  und  $i$  der Gesamtanzahl der Verdaue. Bei jedem Verdau kann der Transposoninsertion eine Menge  $\bar{M}_j$  an möglichen Positionen innerhalb des Genoms zugeordnet werden. Das Fragment mit Transposoninsertion befindet ist in allen  $\bar{M}_j$ . Nach jedem Verdau kann die Anzahl an möglichen Transposoninsertionsstellen auf die Schnittmenge

$$\bigcap_{j \in \{1, \dots, i\}} \bar{M}_j$$

aller Mengen der möglichen Positionen reduziert werden.



**Abbildung 4.3.: Beispiel der Fragmentüberlagerungen.**

Drei Restriktionsverdauungen (a-c) mit jeweils drei unterschiedlichen Enzymen  $i = 3$  werden durchgeführt. Der Kreis symbolisiert die bakterielle DNA  $z = (z_h)$ . Schritt (a) zeigt in Punkt 1. die DNA, welche mit dem ersten Restriktionsenzym  $j = 1$  geschnitten wird. Die roten markierten Fragmente in 2. besitzen ungefähr die zu suchende Länge  $x_1$  und befinden sich somit in der Menge  $\bar{M}_1$ . 3. zeigt die möglichen Stellen der Transposoninsertion  $\bigcap_{j \in \{1\}} \bar{M}_j$ . Das grüne Restriktionsenzym  $j = 2$  schneidet die DNA im 1. Punkt des Schritts (b). Die DNA-Abschnitte  $\bar{M}_2$  mit richtiger Länge  $x_2$  werden in 2. Punkt in grün markiert. Nur bei Überlagerung  $\bigcap_{j \in \{1,2\}} \bar{M}_j$  von rot und grün (3.) kann den Abschnitt der DNA das Transposon beinhalten. In Schritt (c) 1. werden in gelb Schnitte der dritten Restriktionsreaktion  $j = 3$  gezeigt. Die möglichen Fragmente  $\bar{M}_3$  mit der richtigen Länge  $x_3$  werden in Punkt 2. mit gelb markiert. Nur ein mögliches Fragment  $\bigcap_{j \in \{1,2,3\}} \bar{M}_j$ , auf dem das Transposon inseriert sein kann, bleibt im 3. Punkt über.

Mit dieser Fragmentüberlagerungen kann sich die Anzahl möglicher Fragmentposition mit Transposoninsertion minimieren. Es werden so viele Re-



striktionsverdaue genutzt, bis

$$\bigcap_{j \in \{1, \dots, i\}} \bar{M}_j = 1$$

ist. Im Idealfall kann sie auf eins reduziert werden. Ist die Anzahl der möglichen Fragmente  $\bigcap_{j \in \{1, \dots, i\}} \bar{M}_j = 1$ , so konnte das Transposon innerhalb des Genoms kartiert werden. Dieses ist genauer veranschaulicht in Abbildung 4.3.

Innerhalb des *InFiRe*-Algorithmus wird ebenfalls die Übereinstimmung

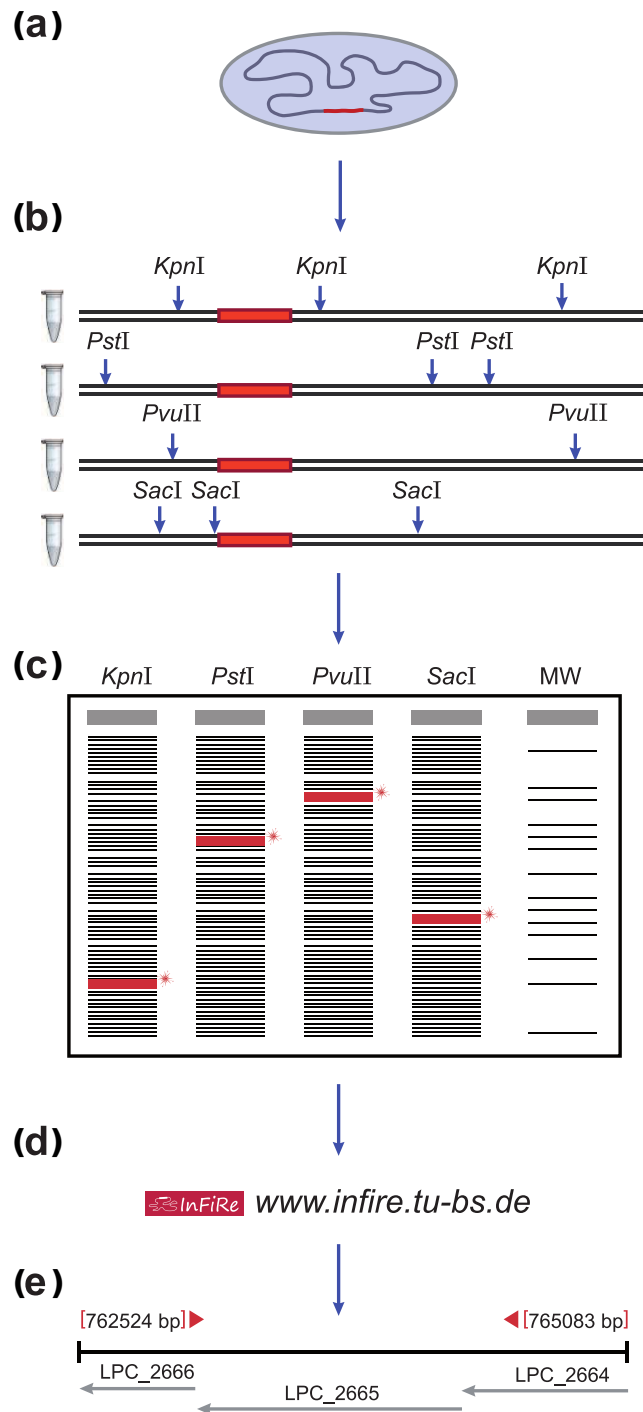
$$U = 100 \cdot \left(1 - \frac{|x - F_k|}{F_k}\right). \quad (4.5)$$

mit der gemessenen Fragmentlängen  $x$  und der kalkulierten Fragmentlängen  $F_k$  berechnet. Diese kann nach dem *InFiRe*-Algorithmus bestimmt werden und wird in Prozent angegeben. Die Übereinstimmung  $U$  zeigt, wie genau die Fragmentlängen von dem Southern-Blot abgelesen wurden. Ist die Abweichung zu stark, sollte die Ablesung des Southern-Blots überprüft werden.

## 4.4. Arbeitsablauf des *InFiRe*-Algorithmus

Der *InFiRe*-Algorithmus basiert auf der Individualität und Reproduzierbarkeit der Restriktionsfragment-Verteilung. Werden mehrere Restriktionsreaktionen mit unterschiedlichen Restriktionsenzymen durchgeführt, wird die Position des Transposons innerhalb eines Genoms kartiert. Der Ablauf der Strategie wird in Abbildung 4.4 dargestellt.

- a) Das zu untersuchende bakterielle Genom  $z = (z_h)$  wird mit einem Transposon mutiert. Die Position des Transposons ist nach der Mutation



**Abbildung 4.4.: Arbeitsablauf zur Detektion von Transposoninsertions-Stellen mittels des Programms *InFiRe*** (a) Eine Transposonmutation wird mit dem zu untersuchenden bakteriellen Genom  $z = (z_h)$  durchgeführt. (b) Restriktionsreaktionen mit  $i$  unterschiedlichen Restriktionsenzymen hydrolysieren die bakterielle DNA. (c) Die Länge  $x$  der Fragmente mit Transposon von jeder Restriktionsreaktion wird durch eine Southern-Blot-Hybridisierung ermittelt. (d + e) Diese Längen lassen eine Berechnung mittels des *InFiRe*-Programmes der Insertionsstelle zu.

nicht bekannt. Voraussetzung für den *InFiRe*-Algorithmus ist die Kenntnis sowohl über die Sequenz des Genoms des Bakteriums als auch über die Sequenz des Transposon. Es muss gewährleistet sein, dass die benutzten Restriktionsenzyme nicht innerhalb des Transposons schneiden, da sonst eine differierende Fragmentlängenverteilung entsteht.

- b) Nach der Transposonmutation wird mit der genomischen DNA mehrmals Restriktionverdaue mit  $i$  unterschiedlichen Enzymen durchgeführt. Die Anzahl an benötigten Restriktionsreaktionen kann durch die in Eigenarbeit erstellten und in dem *InFiRe*-Paket enthaltenen Statistiken berechnet werden (siehe Kapitel 5.1). Jedes Restriktionsenzym hat eine andere Erkennungssequenz und schneidet daher an anderen Stellen im Genom (siehe Kapitel 3.2). Da sich die Schnittstellen pro Restriktionsreaktion unterscheiden, sind die Fragmente auf denen sich das Transposon befindet nach jeder Restriktionsreaktion unterschiedlich lang.
- c) Die Länge der Fragmente  $x$  mit der Transposoninsertion wird mit einer Southern Blot-Hybridisierung (siehe Kapitel 3.3) und einer Transposon-spezifischen Sonde bestimmt. Die Sonde hybridisiert mit dem Transposon und markiert somit das gesuchte Fragment. Die Länge dieses Fragments wird durch einen mitlaufenden Größenstandard geschätzt, wodurch sich eine nur ungefähre Fragmentlänge bestimmen lässt. Daher wird eine Ablesungenauigkeit  $a$  eingeführt. Diese liegt bei ca. 20%.
- d) + e) Mit der Länge des Fragmentes und die Verteilung der Fragmentlängen kann die Insertionsposition des Transposon mittels der *InFiRe*-Software kartiert werden (siehe Abbildung 4.3). Um die *InFiRe*-Software nutzen zu können, werden die Daten der Restriktionsreaktionen benötigt. Diese beinhalten die Sequenz des mutierten Genoms, die Restrik-

tionsreaktionen mit Restriktionsenzymen und dazu passenden Längen der Fragmente mit Transposon, die Ableseungenauigkeit, die Länge des Transposons oder deren exakte Sequenz. Der Algorithmus ist in Kapitel 4.3 erläutert.

Nach der Insertionsstellenbestimmung kann durch eine PCR diese Position bestätigt werden.

## 4.5. R-Pakete

Der *InFiRe*-Algorithmus wurde in Eigenarbeit in zwei R-Paketen mit Namen *Digestion* und *InFiRe* implementiert. Das Paket *Digestion* ermöglicht virtuelle Restriktionsverdaue bekannter Genome. Das *InFiRe*-Paket beinhaltet den vorgestellten Algorithmus. Die Pakete werden mit Anleitung auf der Internetseite <http://www.InFiRe.tu-bs.de/download> zur Verfügung gestellt.

### 4.5.1. *Digestion*-Paket

Das in R erstellte Paket *Digestion* beinhaltet Methoden, welche im Zusammenhang mit Restriktionsreaktionen in bakteriellen Genomen stehen. Bei Kenntnis über die Sequenz des Genoms bietet das *Digestion*-Paket Methoden an, welche das Genom virtuell zerschneiden und darstellen können. Es sind mehrere Restriktionsenzyme pro Restriktionsreaktion, unterschiedlich viele DNA-Sequenzen und mehrere Restriktionsreaktionen gleichzeitig durchführbar.

Der Anwender hat für die Restriktionsreaktionen eine Auswahl an bekannten Enzymen<sup>1</sup>. Es ist auch möglich, weitere noch nicht bekannte Restriktionsenzyme hinzuzufügen. Die Fragmente können direkt angezeigt oder in

---

<sup>1</sup><http://rebase.neb.com/cgi-bin/asymmllist>

einem Histogramm geplottet werden.

#### **4.5.2. *InFiRe*-Paket**

Für den *InFiRe*-Algorithmus wird die Fragmentlängen-Verteilung zu jeder Restriktionsreaktion als Grundvoraussetzung benötigt. Diese wird mit dem *Digestion*-Paket berechnet. Ebenso werden die Längen der Fragmente mit inseriertem Transposon, die Transposonlänge und die Ableseungenauigkeit des Southern Blot benötigt.

Das Transposon inseriert in die DNA und verlängert damit das zu suchende Fragment um seine eigene Länge. Die Länge des Fragments mit Transposon wird durch Southern-Blot-Hybridisierung (siehe Kapitel 3.3) bestimmt. Mit diesen Angaben kann der *InFiRe*-Algorithmus die Position des Transposon und die Übereinstimmungsgenauigkeit  $U$  berechnen.

# 5. Ergebnisse und Diskussion

## 5.1. Statistik

Um festzustellen, wie viele Restriktionsreaktionen die *InFiRe*-Analyse benötigt, wurden in Eigenarbeit zwei statistische Modelle entwickelt. Hierbei wird zwischen einer globalen und der expliziten Statistik unterschieden.

Das globale Modell berechnet die Anzahl an benötigten Restriktionsverdau  $i$  basierend auf der Anzahl der Basenpaare des bakteriellen Genoms  $n$  und der Länge der Erkennungssequenz der Restriktionsenzyme  $\ell$ . Sie kann die benötigte Anzahl an Verdauungen ohne genauere Kenntnis über Genomsequenz, Transposonsequenz oder Restriktionsenzyme berechnen.

Das explizite Modell berechnet im Gegensatz zu der globalen Statistik anhand der exakten Genomsequenz  $z = (z_h)$  und explizit angegebenen Erkennungssequenzen der Restriktionsenzyme  $z^R = (z_h)^R$  die Anzahl der benötigten Reaktionen.

Ein Vergleich der Modelle zeigt, dass beide zu annähernd gleichen Ergebnissen kommen.

### 5.1.1. Globale Statistik

Die globale Statistik berechnet die Anzahl an nötiger Verdauungen  $i$ , ohne die exakte Genomsequenz oder die Erkennungssequenz der Restriktionsenzyme zu kennen. Sie benötigt nur die Sequenzlänge des bakteriellen Genoms  $n$ , die Länge der Erkennungssequenz der Restriktionsenzyme  $\ell$  und die Ableseungenauigkeit  $a$  des Southern Blot. Daraus kann die theoretische Dichtefunktion der Fragmentlängen mit und ohne Ableseungenauigkeit ( $P_1$  und  $P_2$ ) berechnet werden. Wie schon in Kapitel 4.2 dargestellt, wird davon ausgegangen, dass alle Nukleinbasen gleich häufig und zufällig-verteilt innerhalb des Genoms vorkommen.

Die Fragmentlängen-Dichtefunktion  $P_1$  von jeder einzelnen Restriktionsreaktion  $i$  mit der vorgegebenen Länge der Erkennungssequenz der Restriktionsenzyme  $\ell$  muss für die theoretische Statistik berechnet werden. Der Parameter  $x$  steht für die Fragmentgrößen. Die Dichtefunktion der Fragmentgrößen  $P_1$  kann durch Gleichung (4.3) berechnet werden, siehe Satz 2.

Die Verteilung  $P_2$  mit Ableseungenauigkeit wird mit der Dichtefunktion  $P_1$  berechnet. Für eine Fragmentgröße, welche auf dem Intervall  $[x - a \cdot x, x + a \cdot x]$  liegt, wird  $P_2(x) = \int_{x-a \cdot x}^{x+a \cdot x} P_1(t) dt$  definiert und somit ist

$$P_2(x) = \left( e^{-\frac{x-a \cdot x}{m}} - e^{-\frac{x+a \cdot x}{m}} \right), \quad \forall x \geq 0. \quad (5.1)$$

Das Maximum von  $P_2(x)$  ist die maximale mögliche Anzahl an in frage kommender Fragmenten

$$b := \max(P_2(x)), \quad \forall x \geq 0 \quad (5.2)$$

und wird mit der Fragmentlänge  $x$ , der Ableseungenauigkeit  $a$  und der Wahrscheinlichkeitsverteilung  $P_2$  berechnet.

Mit  $b$  kann die ungefähre Anzahl an erforderlichen Restriktionsreaktionen  $i$  berechnet werden. Mit jeder Restriktionsreaktion verringert sich die Anzahl  $z$  an möglichen Genompositionen der Insertion. Die Anzahl  $z$  ist invers proportional zu der Anzahl der Restriktionsreaktionen. Mit jeder hinzukommenden Restriktion wird  $z$  neu berechnet, bis  $z \leq 1$ . Um die Anzahl benötigter Restriktionsreaktionen zu berechnen, benötigt man

- $b$ , die maximale Anzahl an Fragmente für eine gegebene Fragmentgröße,
- $k$ , die Anzahl an Restriktionsenzymen pro Restriktionsreaktion,
- $g$ , die Länge der gesamten Genomsequenz,
- $\ell$ , die Länge der Erkennungssequenz des Restriktionsenzym,
- $n(1)$ , die Anzahl aller Fragmente und
- $c(1)$ , das Verhältnis aller Fragmente mit der Länge  $x$  zu der Anzahl aller Fragmente  $n(1)$ .

Der unten gezeigt Pseudocode berechnet dann  $z = n(j) \cdot c(j)$ , wobei  $i$  die Nummer der Restriktionsreaktionen bezeichnet.

Nach jeder Schleife  $j$  wird die Anzahl möglicher Fragmente  $n(j) \cdot c(j)$  berechnet. Gibt es weniger als ein mögliches Fragment ( $n(j) \cdot c(j) \leq 1$ ), reflektiert die bis zu diesem Punkt ausgeführte Anzahl an Schleifen die Anzahl an nötigen Restriktionsreaktionen ( $j = i$ ). Es ist möglich, mehr Restriktionsreaktionen durchzuführen, als berechnet wurden.



---

```
n(1) = k*(g/4^l)
c(1) = b/n(1)
j = 1
while (n(j)*c(j)>=1)
{
    j      = j+1
    c(j)   = c(j-1)*c(1)
    n(j)   = n(1)+n(j-1)
}
```

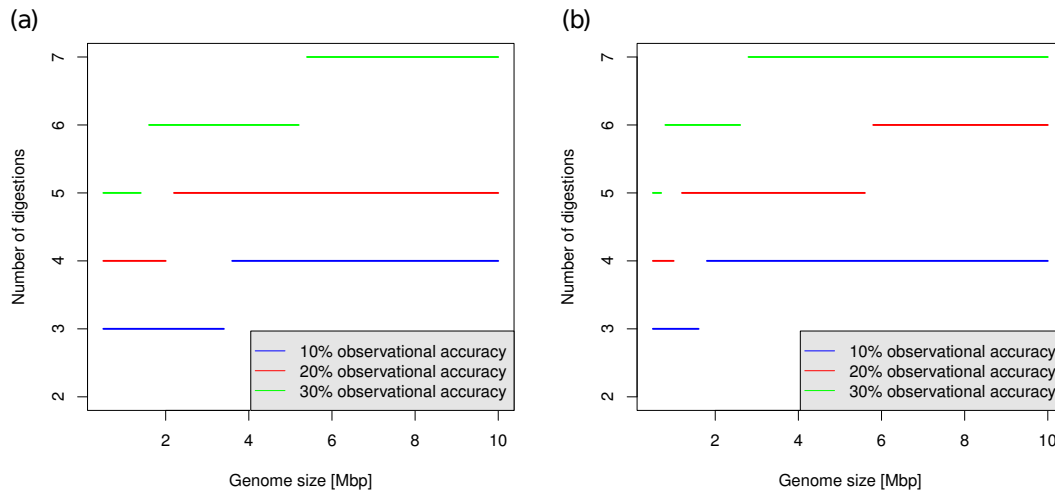
---

In Abbildung 5.1 wird eine vereinfachte Darstellung der Ergebnisse der globalen Statistik dargestellt. Diese findet man innerhalb des Webinterfaces<sup>1</sup>. Die Anzahl  $i$  der für das *InFiRe*-Programm benötigten Restriktionsreaktionen wird mit je einem und je zwei Restriktionsenzymen pro Restriktionsreaktion gezeigt. Wird mit zwei Restriktionsenzym pro Verdau das Genom geschnitten, so verdoppelt sich die Schnittanzahl pro Verdau. Die globale Statistik kann direkt über die R-Pakete ausgeführt werden.

Die globale Statistik berechnet die Anzahl an benötigten Restriktionsreaktionen mit der Anzahl an Basenpaaren des bakteriellen Genoms  $n$ , mit der Länge der Erkennungssequenz der Restriktionsenzyme  $\ell$  und mit der Ablesungenauigkeit  $a$ . In Abbildung 5.1 werden die Ergebnisse der Berechnung der globalen Statistik in einem Diagramm gezeigt. Für diese Darstellung (Abbildung 5.1) wird davon ausgegangen, dass alle anzuwendenden Restriktionsenzyme eine hexamere Erkennungssequenz besitzen, daher ist die Länge der Erkennungssequenz  $\ell = 6$ .

---

<sup>1</sup><http://www.InFiRe.tu-bs.de/estimation>



**Abbildung 5.1.:** Die theoretische Anzahl an benötigten Restriktionsreaktionen für die Analyse mittels *InFiRe*. Hier bei wird unterschieden, ob (a) Reaktionen mit einem Restriktionsenzym und (b) mit zwei Restriktionsenzymen durchgeführt wird. Beide Diagramme zeigen die Anzahl der benötigten Restriktionsreaktionen abhängig von der Länge des Genoms und von den drei verschiedenen Ableseungenauigkeiten (blau 10 %, rot 20 % and grün 30 %) des Southern Blot an. Bei allen wird eine Erkennungssequenz-Länge der Restriktionsenzyme von  $\ell = 6$  angenommen.

Es ist empfehlenswert, eine Restriktionsreaktion mehr durchzuführen, als im Diagramm dargestellt. Die Wahrscheinlichkeit eines exakten Ergebnisses mittels des *InFiRe*-Algorithmus wird dadurch erhöht und die Transposoninsertion kann genauer kartiert werden.

### 5.1.2. Explizite Statistik

Die explizite Statistik berechnet die Anzahl an Restriktionsverdau mit der direkt angegebenen Genomsequenz  $z = (z_h)$  und den möglichen Restriktionsenzymen mit den dazugehörigen Erkennungssequenzen  $z^R = (z_h)^R$ . Im Gegensatz zu der globalen Statistik berechnet sie zusätzlich die günstigste Kombination von Restriktionsenzymen mit einer gewissen Fehlerwahr-

scheinlichkeit. Die Fehlerwahrscheinlichkeit zeigt die Wahrscheinlichkeit an, mehr Restriktionsenzyme zu benötigen, um ein mögliches Fragment als Ergebnis zu erhalten.

Aus den Daten kann für jede Restriktionsreaktion  $i$  direkt die exakte Fragmentlängenverteilung mit und ohne Ableseungenauigkeit berechnet werden.

**Definition 2.** Die Verteilung der Fragmentlängen aus angegebenen Daten nach Restriktionsverdau  $i$  wird angegeben durch  $P_1^i$ . Die Verteilung der Fragmentlängen mit einer Ableseungenauigkeit nach Restriktionsverdau  $i$  wird angegeben durch  $P_2^i$ .

**Definition 3.** Die maximale Anzahl an passenden Fragmenten

$$b_i := \max(P_2^i(x)) = P_2^i(x_i^{\max}), \quad \forall x \geq 0 \quad (5.3)$$

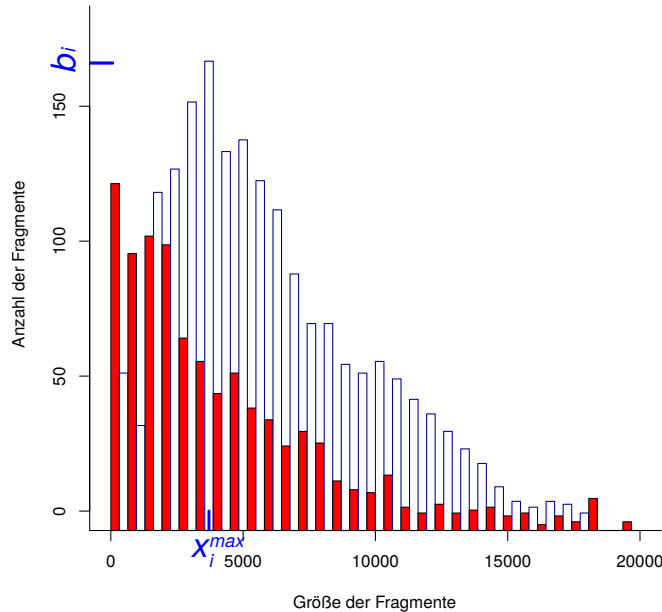
des Verdau  $i$  kann mit der Fragmentlängenverteilung  $P_2^i$  mit der Ableseungenauigkeit  $a$  und der dazugehörigen Fragmentlänge  $x_i^{\max}$  bestimmt werden (siehe Abbildung 5.2).

Um die maximale Anzahl der erforderlichen Restriktionsreaktionen zu berechnen, werden ausschließlich Fragmente mit einer Fragmentlänge, welche innerhalb eines Intervalls  $[x_i^{\max} - a \cdot x_i^{\max}, x_i^{\max} + a \cdot x_i^{\max}]$  liegt, betrachtet. Alle Fragmente mit einer Länge, die auf diesem Intervall ist, werden der Menge  $Q_i$  zugeordnet. Die maximale Anzahl an Fragmenten kann für jeden Restriktionsverdau  $i$  bestimmt werden und die dazugehörige Länge  $x_i^{\max}$  ausgegeben werden. Das Intervall betrachtet alle Fragmente um diese Länge plus/minus der Ableseungenauigkeit  $a$  herum.

**Satz 3.** Für jede Restriktionsreaktion  $i$  befinden sich maximal viele Fragmente auf dem Intervall

$$Q_i = \bar{M}(x_i^{\max}) = \int_{x_i^{\max} - a \cdot x_i^{\max}}^{x_i^{\max} + a \cdot x_i^{\max}} L^{-1}(t) dt \quad (5.4)$$

bei einer Fragmentgröße  $x_i^{\max}$  (siehe Kapitel 4.2).



**Abbildung 5.2.:** Ein Beispiel für die Fragmentlängenverteilung. Das *L. pneumophila* Corby wurde hierbei mit der Restriktionsenzyme *EcoRI* geschnitten (rote Balken). Mit einer Ableseungenauigkeit von 20% wird eine Verteilung von Fragmenten zu einer Links-bündigen Verteilung (blaue Balken). Das Maximum der Fragmente ist gekennzeichnet mit  $b_i$  und einer Fragmentlänge von  $x_i^{\max}$ .

In jedem Restriktionsverdau  $i$  wird die Menge  $Q_i$  für das Intervall  $[x_i^{\max} - a \cdot x_i^{\max}, x_i^{\max} + a \cdot x_i^{\max}]$  berechnet. Alle Fragmente in  $Q_i$  werden als in Frage kommende Fragmente bewertet. Mögliche Fragmente von allen Restriktionsreaktionen können sich überlappen. Die Überlappungen von  $Q_i$  werden in der Variable  $O_i$  gespeichert,  $O_i = \bigcap Q_j$  mit  $j \in [1, 2, \dots, i]$ . Der Parameter  $j$  beinhaltet alle bis zu diesem Zeitpunkt durchgerechnete Restriktionsreaktionen  $i$ . Die Anzahl der Restriktionsreaktionen wird weiter erhöht, bis sich nur noch eine Überlappung ergibt (siehe Abbildung 4.3), also  $\#O_i = 1$ .

Das Resultat dieser Berechnung wird mit einer Fehlerwahrscheinlichkeit  $e$  bewertet. Diese berechnet sich aus der Anzahl an Überlappungen aller bis

dahin ausgeführter Verdauung ( $o_i = \#O_i$ ), der Anzahl aller Fragmente  $g_i$  pro Restriktionsreaktion  $i$  und der Anzahl an in Frage kommenden Fragmenten  $f_i = \# Q_i$  im Verdau  $i$ .

$$e(i) := \prod_{j=1}^{i-1} \left( \frac{g_j - o_j}{g_j} \right)^{f_{(j+1)}} = \prod_{j=1}^{i-1} \left( 1 - \frac{o_j}{g_j} \right)^{f_{(j+1)}} \quad (5.5)$$

The image shows two panels of a web interface. Panel (a) is titled 'Select Genome and Replicons' and contains a 'Genomes' dropdown menu with 'Legionella pneumophila Corby' selected, an 'Add ->' button, a '<- Remove' button, a 'Replicons' list with 'Legionella pneumophila str. Corby chromosome', an 'Upload Genome File (gbk)' section with a 'Durchsuchen...' button and 'Keine Datei ausgewählt.' text, and an 'Upload Genome' button. Panel (b) is titled 'Digestion Parameters' and contains a 'Choose at least 5 or more different restriction enzymes.' instruction. It has three columns: 'Enzymes' with a list of enzymes (AalI, AagI, AarI, AasI, AatI, AatII, AauI, AbaI, AbeI, Abri), a 'Selection' column with an empty list, and a 'Digestion' column with a list of enzymes (EcoRI, SacI, ScaI, KpnI, PstI, SphI). There are 'Add ->' and '<- Remove' buttons between the columns. To the right is an 'Observational Accuracy' input field with '20' and a '%' sign. At the bottom are 'Statistics' and 'Reset' buttons.

**Abbildung 5.3.: Stufen der expliziten Statistik des Webinterfaces**

(a) Die Auswahl des bakteriellen Genoms und deren Replikons. Hier wurde das Chromosom von *Legionella pneumophila* Corby gewählt

(b) Eingabe der möglichen Restriktionsreaktionen mit möglicher unterschiedlicher Anzahl an Restriktionsenzymen und der Ableseungenauigkeit (20%) des folgenden Southern Blots. Hier sind sechs Restriktionsenzyme gewählt: *EcoRI*, *SacI*, *ScaI*, *KpnI*, *PstI* und *SphI*.

Hierbei ist  $(g_j - o_j)/g_j = 1 - o_j/g_j$  der Anteil des Genoms ohne Fragmenten, die in Frage kommen. Mit jedem Fragment des nächsten Verdaus  $f_{j+1}$  kann dieser Anteil *falsch* getroffen werden.

Die explizite Statistik ist direkt über das Webinterface<sup>2</sup> oder über die R-Pakete ausführbar. Die explizite Statistik ist innerhalb des Webinterfaces in zwei Schritte unterteilt (siehe Abbildung 5.3).

Zunächst muss das zu erforschende bakterielle Genom festgelegt werden (siehe Abbildung 5.3 a)). Es kann auf alle sequenzierten bakteriellen Genome der NCBI-Webseite zugegriffen werden (siehe Kapitel 2.2.1). Bakterielle Genome können mehrere Replikons enthalten (zum Beispiel Chromosomen und Plasmide). Wird ein bakterielles Chromosom ausgewählt, so werden direkt alle bekannten Replikons mit angegeben. Soll eines dieser Replikons nicht mit analysiert werden, muss dieses aktiv zurückgesetzt werden.

Result	
Minimum number of digestions	4
Optimal enzymes	KpnI, SacI, ScaI, SphI
Error probability	1.006e-11 %

#### Abbildung 5.4.: Resultat der *InFiRe*-Statistik

Hier wird das Resultat der in Abbildung 5.3 gezeigten Statistik dargestellt. Die minimale Anzahl an empfohlenen Restriktionsreaktionen ist hier mit vier angegeben. Die zu nutzenden Restriktionsreaktionen sollen mit den folgenden Restriktionsenzymen durchgeführt werden: *KpnI*, *SacI*, *ScaI* und *SphI*. Die Fehlerwahrscheinlichkeit wurde mit  $10^{-11}$  angegeben.

Alle zur Verfügung stehenden Restriktionsenzyme werden in gewünschten Gruppierungen für jede der Restriktionsreaktionen angegeben (siehe Abbildung 5.3 (b)). Das explizite Statistikprogramm wird nur dann ausgeführt,

<sup>2</sup><http://www.infire.tu-bs.de/statistics>

wenn mindestens fünf verschiedene Restriktionsreaktionen angegeben werden. Je mehr angegeben werden, desto besser kann die Kombination der Restriktionsreaktionen durch das Programm berechnet werden. Die Berechnung dauert nur einige Sekunden.

Das Ergebnis der expliziten Statistik beinhaltet nicht nur die Anzahl an benötigten Restriktionsreaktionen, sondern eine Liste mit den Restriktionsreaktionen, die genutzt werden sollen, um ein optimales Ergebnis zu erzielen. In Abbildung 5.4 wird das Ergebnis der Expliziten Statistik dargestellt. Weniger Restriktionsreaktionen durchzuführen, ist nicht ratsam, da dadurch kein eindeutiger Ort für die Transposoninsertion bestimmt werden kann.

### 5.1.3. Vergleich der Statistiken

Die explizite und globale Statistik wurden mit einer zufälligen Simulation verglichen. Die explizite Statistik wurde 360 mal durchgeführt. Hierbei wurden zehn verschiedenen Listen von zufällig ausgewählten Restriktionsverdauen, mit je einem oder zwei Restriktionsenzymen (siehe Tabelle 5.1), 12 verschiedenen bakteriellen Genomen und drei verschiedenen Ableseungenauigkeiten (10 %, 20 % und 30 %) genutzt.

Die explizite Statistik gab im Schnitt eine Anzahl von  $\mu_e = 4.255$  benötigten Restriktionsreaktionen mit einer Standardabweichung von  $sd_e = 1.139$  an. Die globale Statistik berechnete hingegen eine durchschnittliche Anzahl an Reaktionen von  $\mu_g = 4.833$  mit einer Standardabweichung von  $sd_g = 1.221$ . Im Schnitt gibt die explizite Statistik daher 0.578 Reaktionen mehr an als der globale Ansatz.

**Tabelle 5.1.:** Die zehn verschiedenen Restriktionsreaktions-Listen mit je sieben Restriktionsreaktionen. Fünf davon beinhalten nur Einzel-Restriktionsreaktionen, fünf sind Doppel-Restriktionsreaktionen. Die Enzyme sind zufällig ausgewählt.

	Restriktionsenzyme
1	<i>AaaI, BcoAI, SurI, HgiDII, EcoRI, SpaHI, EgeI</i>
2	<i>PmaCI, SsrI, ScaI, BamHI, AccBSI, MluB2I, SpmI</i>
3	<i>KpnI, SacI, SplI, ZrmI, TliI, EcoRV, MluI</i>
4	<i>SacI, NunII, Sru4DI, NaeI, SacII, SruI, HalII</i>
5	<i>NsbI, FauNDI, LlaG2I, Kpn2I, ScaI, PfaAIII, EcoRI</i>
6	<i>PmaCI:RtrI, SsrI:Pfl8I, NsiCI:SpeI, BamHI:Mlu23I, Mlu3II:AccBSI, Pfl23II:MluB2I, SpmI:NbII</i>
7	<i>KpnI:SsoI, NspSAIV:NarI, SacI:SplI, Rtr63I:ZrmI, MluNI:TliI, EcoRV:SphI, PfaAII:MluI</i>
8	<i>NsbI:YenI, PaePI:FauNDI, LlaCI:LlaG2I, Kpn2I:SphI, SstI:ScaI, PfaAIII, Pfl2II:NsiI</i>
9	<i>SacI:XmaJI, PleI9I:NunII, PaeR7I:Sru4DI, NaeI:CsiAI, MxaI:SacII, PaeQI:SruI, HalII:EcoT22I</i>
10	<i>AaaI:AbrI, BcoAI:BspI19I, SurI:SspCI, HgiDII:DpaI, EcoRI:CscI, XcaI:SpaHI, EgeI:HjaI</i>

Die Abweichung wurde mit einem Mittelwertsvergleich, einem t-Test, bestätigt. Der t-Test ergab, dass mit einem p-Wert von  $p = 2.2 \cdot 10^{-16}$  die explizite Statistik kleiner ist, als die globale Statistik. Das Resultat zeigt, dass der globale Ansatz die Anzahl an maximal benötigten Restriktionsreaktionen nur leicht überschätzt. Dieses stellt in der Praxis kein Problem dar, da es immer vorteilhaft ist, wenn mehr Reaktionen durchgeführt werden und höchstens ein genaueres Ergebnis dadurch entsteht.

## 5.2. Analyse innerhalb des Webinterface

Um die Nutzung des *InFiRe*-Programms zu verbessern, wurden die Pakete auf der Webseite <http://www.InFiRe.tu-bs.de> zur Verfügung gestellt. Es stehen dort sowohl die R-Pakete und deren Anleitungen als Download, als auch die *InFiRe*-Analyse zur direkten Anwendung als Webinterface zur Verfügung.



Bevor die *InFiRe*-Analyse gestartet wird, muss mittels der schon berechneten globalen oder der im Webinterface zur Verfügung gestellten expliziten Statistik die Anzahl an benötigten Restriktionsreaktionen für den Versuchsaufbau berechnet werden. Diese Methoden wurden in Kapitel 5.1 etabliert.

Das Webinterface zur *InFiRe*-Analyse (<http://www.InFiRe.tu-bs.de/analyze>) ist in drei Teile unterteilt, die Auswahl des bakteriellen Genoms und seinen Replikons, die Transposonwahl und die Eingabe der Restriktionsreaktions-Daten (siehe Abbildung 5.5).

Bereits sequenzierte bakterielle Genome und Replikons, welche auf der Webseite des NCBI (siehe Kapitel 2.2.1) verfügbar sind, können im ersten Schritt ausgewählt werden (Abbildung 5.5 (a)). Die gleichen Genome sind ebenfalls in der expliziten Statistik verwendbar.

Um den *InFiRe*-Algorithmus auszuführen, muss die Sequenz des inserierten Transposons eingegeben werden (Abbildung 5.5 (b)). Hierzu stehen drei Eingabemethoden bereit. Die Transposon-Sequenz kann aus einer vorhandenen Liste von gegebenen Transposons ausgewählt werden. Ist das gewünschte Transposon nicht in dieser Liste zu finden, bietet das Webinterface zwei weitere Möglichkeiten an. Zum einen ist es möglich, eine die Transposon-Sequenz enthaltene FASTA-Datei (siehe Kapitel 2.5.2) zu erstellen und diese dann hochzuladen, so dass das *InFiRe*-Programm darauf zugreifen kann. Zum anderen kann die Transposon-Sequenz auch über ein Textfeld eingegeben werden. Hierbei muss die Sequenz ebenfalls den Anforderungen des FASTA-Formats entsprechen. Wird nach der Eingabe der Transposon-Sequenz der *Select*-Knopf gedrückt, werden die Restriktionsenzyme angezeigt, welche nicht in das genutzte Transposon schneiden (siehe Abbildung 5.5). Diese Enzyme werden dann von dem *InFiRe*-Algorithmus in der Analyse unterstützt.

The figure consists of three panels, (a), (b), and (c), showing different stages of the InFiRe web interface.

**(a) Select Genome and Replicons:** This panel has a title bar 'Select Genome and Replicons'. It contains a 'Genomes' section with a dropdown menu showing 'Legionella pneumophila Corby'. Below it are 'Add ->' and '<- Remove' buttons. The 'Replicons' section has a list box showing 'Legionella pneumophila str. Corby chromosome'. At the bottom, there is an 'Upload Genome File (gbk)' section with a 'Choose File' button (showing 'No file chosen') and an 'Upload Genome' button.

**(b) Select Transposon:** This panel has a title bar 'Select Transposon'. It contains a 'Transposon Source' section with a list box showing 'List of Transposons', 'Upload FASTA File', and 'Paste Sequence (FASTA)', with a 'Select' button below. The 'Select Transposon' section has a list box showing several transposons: 'Tn5(TET-1)', 'Tn5(TN-TET-1)', 'Tn5(blaM-R6Kyor)', 'Tn5(oriV-KAN-2)', 'Tn5(p15Aori-KAN-2-T7Exp)', and 'miniTn10(Kan)'.

**(c) Digestion Parameters:** This panel has a title bar 'Digestion Parameters'. It contains four main sections: 'Enzymes' (a list box with various restriction enzymes like SfuI, SgrAI, etc.), 'Selection' (a list box with 'EcoRI' and 'SphI'), 'Digestion' (a list box with enzyme pairs like 'EcoRI:KpnI', etc.), and 'Fragment Sizes' (a list box with sizes like 5100, 5000, etc.). There are 'Add ->' and '<- Remove' buttons between the sections. At the bottom, there is an 'Observational Accuracy' section with a text input '15' and a '%' sign, and a 'Fragment' section with a text input '6000' and a 'bp' unit. There are also '>- Add' and '<- Remove' buttons. At the very bottom are 'Analyze' and 'Reset' buttons.

**Abbildung 5.5.: Drei verschiedene Stufen der Analyse des *InFiRe*-Webinterfaces**

(a) Die Auswahl des bakteriellen Genoms und deren Replikons. (b) Die Auswahl des Transposons. (c) Die Restriktionsreaktions-Daten.

Im dritten Schritt müssen die Restriktionsreaktionen und die dazugehörigen Ergebnisse aus dem Southern Blot definiert bzw. eingegeben werden (siehe Abbildung 5.5 (c)). Aus der berechneten Liste von Restriktionsenzymen werden die genutzten Restriktionsenzyme passend zu jeder Restriktionsreaktion gewählt. Für jede dieser Reaktionen muss mittels Southern Blot Hybridisierung (siehe Kapitel 3.3) die Länge des Fragments ermittelt werden,

welches das Transposon enthält. Diese Länge muss passend zu der entsprechenden Restriktionsreaktion definiert werden. Die Ableseungenauigkeit  $a$  des Southern Blot wird ebenfalls für den *InFiRe*-Algorithmus benötigt und sollte um 20 % liegen, je nachdem, wie gut der Southern Blot abgelesen werden kann.

Das Programm benötigt nur einige Sekunden, um das Ergebnis zu berechnen. Eine Liste wird ausgegeben (siehe Abbildung 5.7). Diese beinhaltet die Intervalle von Basen, in denen das Transposon inseriert sein kann. Bei genügend Restriktionsreaktionen wird ein Basenintervall ausgegeben, auf welchem sich die Transposoninsertion befindet. Bei zu wenigen Restriktionsreaktionen wird mehr als ein Basenintervall ausgegeben. Das Transposon wird damit nicht auf ein Intervall kartiert, sondern kann sich auf allen angegebenen befinden. Die Liste beinhaltet zu jedem Intervall verschiedene Daten. Angegeben wird das Replikon, welches das Transposon beinhaltet, sowie der Start, das Ende und die Länge des Intervalls, auf dem das Transposon inseriert sein kann.

Wird mehr als ein Ergebnis in der Liste angezeigt, ist die Liste nach einem Übereinstimmungswert  $U$  sortiert (siehe Gleichung (4.5)). Dieser Wert wird aus der Angabe der geschätzten Fragmentlängen und den berechneten Fragmentlängen kalkuliert. Er zeigt daher an, wie zuverlässig die geschätzten Daten sind, und welche Wahrscheinlichkeit das Ergebnis hat, zu den angegebenen Daten zu gehören.

Zu jedem Ergebnis wird ein Link zu der Internetseite des NCBI *map viewer* (siehe Kapitel 2.2.1) angegeben. Der *NCBI map viewer* (Wolfsberg, 2011) zeigt das entsprechende Fragment innerhalb des bakteriellen Genoms mit vielen zusätzlichen Informationen an.

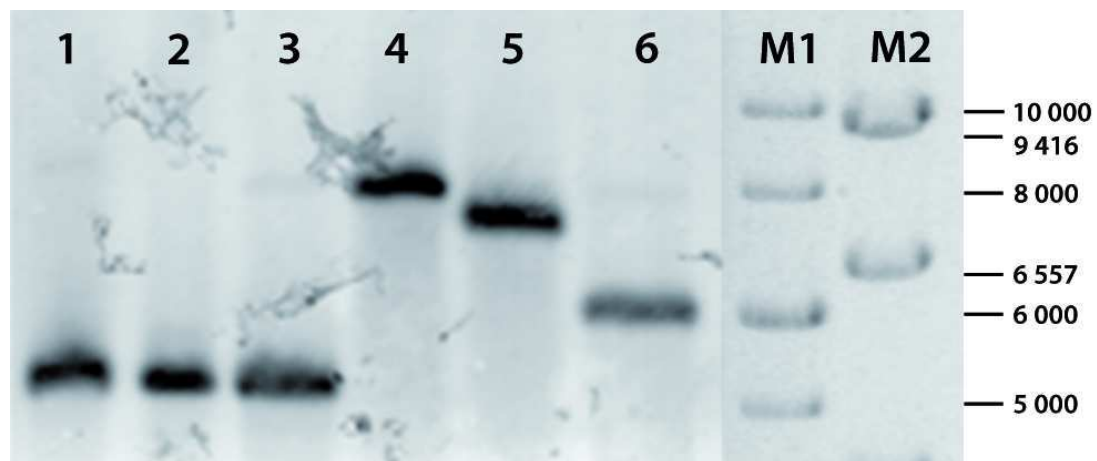
Um das Ergebnis des *InFiRe*-Algorithmus zu bestätigen, kann eine PCR mit zwei definierten Primern durchgeführt werden. Dabei muss entweder

ein Primer eine Bindestelle innerhalb und ein Primer eine Bindestelle außerhalb des Transposon haben, oder beide Primer müssen für die Bestätigung außerhalb des Transposons ihre Bindestelle haben. Die Primer können z. B. mittels des Programms Primer-BLAST (Sayers et al., 2011) erstellt werden.

## 5.3. Fallstudie

Die *InFiRe*-Software wurde erfolgreich mit einer Fallstudie bestätigt. Es wurde eine *L. pneumophila* Corby *mini-Tn10* Transposon-Bibliothek genutzt. Mutanten mit einer abgeschwächten intrazellulären Überlebensrate in Wirtszellen sollten betrachtet werden (Shevchuk and Steinert, 2009). Bei all diesen Mutanten war die Insertionsposition des Transposons unbekannt. Mittels des *InFiRe*-Programms konnte die Transposon-Insertions-Stelle kartiert werden.

Die zur Verfügung stehenden Restriktionsreaktionen sind *EcoRI* mit *SacI*, *EcoRI* mit *ScaI*, *EcoRI* mit *KpnI*, *EcoRI* mit *PstI*, *EcoRI* mit *PvuII* und *EcoRI* mit *SphI*. Mittels der expliziten Statistik wurde berechnet, dass fünf dieser Restriktionsverdau benötigt werden, um ein eindeutiges Ergebnis zu erzielen. Mit der globalen Statistik kommt man ebenfalls auf das Ergebnis von fünf Restriktionsreaktionen mit je zwei Restriktionsenzymen. Durch die Größe des Genoms von *L. pneumophila* Corby (3576470 bp) und der Ableseungenauigkeit von  $a = 20\%$  lässt sich dies berechnen (siehe Abbildung 5.1). Genutzt wurden alle sechs Reaktionen, um die Transposoninsertionsstelle genauer zu kartieren. Jede Reaktion wurde mit zwei Restriktionsenzymen durchgeführt, da die Position des Transposons durch die erhöhte Schnittzahl bei zwei genutzten Restriktionsenzymen pro Verdau genauer bestimmt werden kann.



**Abbildung 5.6.: Southern Blot-Hybridisierung nach sechs Restriktionsreaktionen von *L. pneumophila* Corby.** Angegebene Größen ist in Basenpaaren.

Nach der Mutation von *L. pneumophila* Corby mit dem Transposon *mini-Tn10* wurden sechs Restriktionsverdaue mit der chromosomalen DNA durchgeführt. Zu sehen ist jeweils die Länge des Fragmentes mit Transposoninsertion nachdem *L. pneumophila* Corby::*Tn10* mit den jeweiligen Restriktionsenzymen geschnitten wurde: 1: *EcoRI/KpnI*-HF; 2: *EcoRI/PstI*; 3: *EcoRI/PvuII*-HF; 4: *EcoRI/SacI*-HF; 5: *EcoRI/ScaI*-HF; 6: *EcoRI/SphI*-HF; In der siebten und achten Spur befinden sich Größenstandard ( M1: M1-1kb DNA Ladder SM0311 (Fermentas), M2: M2- Lambda DNA/*HindIII* Marker SM0101 (Fermentas)). Das Fragment mit Transposon wurde durch eine Hybridisierung mit einer DIG-markierten *mini-10-Tn*-spezifischen Sonde sichtbar gemacht.

Die Länge der Fragmente mit Transposon nach den Restriktionsreaktionen wurde mit einer Southern Blot Hybridisierung (Southern, 1975) ermittelt (siehe Abbildung 5.6). Je exakter die Länge des Fragments mit der Southern Blot Hybridisierung bestimmt werden kann, also je kleiner die Ableseungenauigkeit  $a$  ist, desto exakter kann die Position des Transposon bestimmt werden.

Mittels der Daten der Southern Blot Hybridisierung kann nun folgendes innerhalb des Webinterfaces eingegeben werden (siehe Abbildung 5.5):

1. Genutzt wird *L. pneumophila* Corby

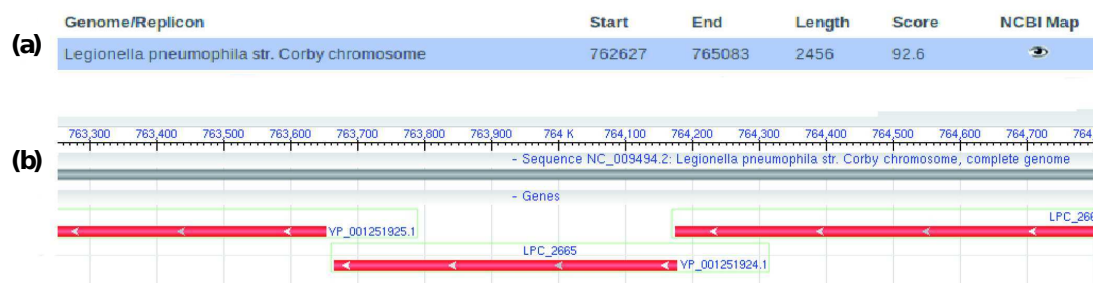
2. Das Transposon *mini-Tn10* ist in der Liste der Transposon verfügbar.
3. Die Daten der Southern Blot Hybridisierung (siehe Tabelle 5.2)

### Tabelle 5.2.: Fragmentlängen nach Restriktionsreaktion

Angegeben sind die Restriktionsenzyme zu jeder Restriktionsreaktion von *L. pneumophila* Corby::*Tn10* und die abgelesenen Fragmentlängen mit inseriertem Transposon.

Restriktionsreaktion	Fragmentlängen
EcoRI und KpnI	5100 bp
EcoRI und PstI	5000 bp
EcoRI und PvuII	5100 bp
EcoRI und SacI	8100 bp
EcoRI und ScaI	7500 bp
EcoRI und SphI	6000 bp

Die Analyse von *InFiRe* mit den Daten aus der Tabelle 5.2 ergab, dass das Transposon auf dem *L. pneumophila* Corby Chromosom zwischen dem Basenpaar 762627 und 765083 liegt, also auf einer Länge von 2456 Basenpaaren.



**Abbildung 5.7.:** Das Resultat des *InFiRe*-Algorithmus (a) Hier wird eine Liste ausgegeben, die alle potentiellen Orte des Transposons beinhaltet. Nutzt man genügend Restriktionsreaktionen, wird nur ein Ort angezeigt. (b) Visualisierung des Ergebnisses mittels des *NCBI Sequence Viewer*

Die berechneten Daten stimmt zu 92.6 % mit den eingegebenen Daten überein. Dies bedeutet, dass der Southern Blot ein genaues Ergebnis liefert und recht exakt abgelesen wurde. Dies wird mit der Gleichung (4.5) berechnet. Übereinstimmungen unter 80 % müssen noch einmal verifiziert werden.

Das *InFiRe*-Programm hat die Position des inserierten Transposon innerhalb des *L. pneumophila* Corby Chromosom auf einen kurzen Abschnitt eingegrenzt. Das Ergebnis (Abbildung 5.7) wurde mittels PCR bestätigt und kann mittels des *NCBI Sequence Viewer* im Genom betrachtet werden.

## 5.4. *InFiRe* im Vergleich zu bisherigen Verfahren

Innerhalb der funktionellen Genomik ist die Transposonmutagenese eine etablierte Methode, um Genen eine Funktion zuzuordnen zu können. Es werden innerhalb der Transposonmutagenese Transposoninsertionen und dessen Auswirkung auf den Phänotyp untersucht. Durch die zufällige Insertion eines Transposon in das Genom kann es zu einer Inaktivierung von einem oder mehreren Genen kommen. Ist bekannt, welches Gen durch die Transposoninsertion inaktiviert wurde, kann bei Veränderung des Phänotyps, dem Gen eine Funktion zugeordnet werden. Der zeit- und kostenintensivste Schritt dieser Methode ist die Lokalisierung der Transposoninsertion. Die Transposoninsertionsstelle lässt sich mit *InFiRe* schnell und zuverlässig herausfinden.

Innerhalb einer Fallstudie (siehe Kapitel 5.3) wurde gezeigt, dass das *InFiRe*-Programm die Position einer Transposoninsertion berechnen kann. Eine Transposonmutante von *L. pneumophila* Corby mit dem *mini-Tn10* Transposon mit unbekannter Insertionsstelle wurde untersucht. Mittels *InFiRe* konnte

die Transposoninsertion kartiert werden. Die Insertion wurde mittels PCR bestätigt. Die Fallstudie zeigt die Funktionalität des *InFiRe*-Programms.

Das *InFiRe*-Programm basiert auf einem Algorithmus, welcher auf der Verteilung von Restriktionsschnittstellen basiert und auf jedem bereits sequenzierten, bakteriellen Genom anwendbar ist. Das Programm stellt eine schnelle, einfach zu handhabende und vor allem kostengünstige Alternative zu bereits vorhandenen Möglichkeiten dar. Eine etablierte Methode ist zum Beispiel die Inverse-PCR (Ochman et al., 1988, Picardeau, 2010), welche Primer nutzt, die innerhalb des Transposons liegen. Hier können jedoch praktische Probleme auftreten. Bei Amplifikationen von Genomabschnitten mit besonders vielen Nukleinbasen Guanin und Cytosin, genannt GC-reiche Genomabschnitte, kann es zu Sekundärstrukturen wie Haarnadeln kommen. Auch eine höhere Schmelztemperatur kann die eine normale PCR-Amplifikation bei GC-reichen Sequenzen aufweisen (Frey et al., 2008). So kann es zum Beispiel möglich sein, dass das Transposon gar nicht gefunden werden kann.

Mehrere Insertionsmutanten können mittels *InFiRe* gleichzeitig analysiert werden. Innerhalb des Labors können somit Transposonmutanten nebeneinander per Restriktionsenzym verdaut werden und per Southern-Blot-Hybridisierung untersucht werden.

Mittels des *InFiRe*-Programms ist es möglich, die Transposoninsertionssuche innerhalb von zwei bis drei Tagen erfolgreich abzuschließen. Bei herkömmlichen Methoden wird zur Insertionsidentifikation mindestens eine Woche benötigt.



## 5.5. Ausblick

Das Software-Programm *InFiRe* kann zuverlässig die Position einer Transposonsinsertionsstelle kartieren. Durch *InFiRe* könnten in Zukunft im Vergleich zu bisherigen Methoden Kosten und Zeit reduziert werden.

*InFiRe* ist in der Lage, Transposons zu detektieren, die in sich wiederholenden Sequenzen eingebettet sind. Diese repetierenden Sequenzen wurden zwar in Genom-Sequenzen kommentiert, jedoch nicht näher auf ihre biologische Rolle untersucht (Hahn, 2009, Hill, 1999, Romero et al., 1999). Mittels des *InFiRe*-Programms können diese repetierenden Sequenzen genauer untersucht werden und möglicherweise einer Rolle zugeordnet werden.

Durch *InFiRe* kann in Zukunft eine Transposon-Mutantenbank einfacher untersucht werden. Es ist möglich, dass der Algorithmus zu einem Hochdurchsatzverfahren erweitert wird. Jede Mutante eines bakteriellen Genoms mit einem Transposon kann mit dem gleichen Set von Restriktionsreaktionen untersucht werden. Mehrere Transposonmutanten, wie in einem Hochdurchsatzverfahren, werden nebeneinander durch Restriktionsenzyme geschnitten und danach mit der gleichen Transposon-spezifischen Sonde im Southern Blot hybridisieren. Hierdurch können Transposon-Mutantenbanken wesentlich einfacher untersucht und Rahmen der funktionellen Genomik eingesetzt werden.



## **Teil II.**

# **Modellierung und Analyse von einer genregulatorischen Feedforward-Schleife mit basaler Expression des zweiten Regulators**



## 6. Einleitung

Bakterien leben in einer sich ständig wechselnden Umwelt. Viele von ihnen haben Mechanismen entwickelt, um sich an Veränderungen anzupassen. Externe Signale und interne Regulatoren bilden dabei zusammen genregulatorische Netzwerke. Mittels Differentialgleichungsmodellen können biologische Systeme auf ihre Stabilität und Robustheit untersucht werden. Theoretische mathematische Modelle können Vorhersagen treffen und Labor-Experimente ergänzen. Teilnetze von Regulationssystemen werden als Netzwerkmodule bezeichnet (Shen-Orr et al., 2002, Alon, 2007). Ein Beispiel dafür ist die sogenannte Feedforward-Schleife (engl. feedforward loop - FFL).

Eine FFL besteht aus zwei Regulatoren, welche von Signalen aktiviert bzw. reprimiert werden und einem Zielprotein (bzw. einer Gruppe von Zielproteinen). Bei der sogenannten Typ-1 kohärente FFL sind alle regulativen Interaktionen Aktivierungen. Aktivieren die Signale die Regulatoren, kommt es nach einer Verzögerungszeit zur Produktion der Zielproteine. Somit können auch kurze Signale gefiltert werden (Mangan and Alon, 2003, Kalir et al., 2005, Wall et al., 2005). Bereits in vielen Systemen konnten FFL-regulierte Gene nachgewiesen werden, so zum Beispiel in verschiedenen Zuckernutzungen von *Escherichia coli* (Mangan et al., 2003, Mangan et al., 2006, Görke and Stülke, 2008).

In dieser Arbeit wird ein Typ-1 kohärentes FFL-System in Bakterien be-

obachtet, bei dem der zweite Regulator der Regulationsschleife dauerhaft exprimiert wird (Schreiber et al., 2006). Mittels eines möglichst einfachen Modell soll untersucht werden, inwiefern die Art der Steuerung nützlich für das Bakterium sein kann und somit vielleicht Rückschlüsse auf das Verhalten des Bakterium gezogen werden können. Die Nitratatmungs-Regulation in *Pseudomonas aeruginosa* wird durch eine solche FFL gesteuert. Es wurde zunächst ein möglichst einfaches Modell dieser Regulation erstellt.

## 6.1. *Pseudomonas aeruginosa*

*Pseudomonas aeruginosa* ist ein Gram-negatives Stäbchenbakterium, und gehört zu der  $\gamma$ -Proteobacteria-Gruppe. Das ungefähr  $2 - 4\mu\text{m}$  langes Stäbchen hat ein relative großes Genom von 6.26 Mbp. Zur Fortbewegung besitzt *P. aeruginosa* Flagellen. Mit seinen Pili kann sich *P. aeruginosa* an Oberflächen festsetzen. Sowohl in Wasser, im Boden, als auch auf der Hautflora des Menschen kann *P. aeruginosa* gefunden werden (Holloway, 1969). Es ist fakultativ anaerob, kann daher sowohl über Oxidation, Denitrifikation aber auch durch Fermentation Energie gewinnen.

*P. aeruginosa* ist ein opportunistisches Humanpathogen. Bei ungefähr 10 % aller Krankenhausinfektionen in Deutschland handelt es sich um eine Infektion mit *P. aeruginosa*. Problematisch sind Infektionen von *P. aeruginosa* (meist im Utrakt, in den Atemwegen oder auf der Haut), da es natürliche Resistenzen oder gar Multiresistenzen gegen Antibiotika aufweisen kann (Morita et al., 2014). *P. aeruginosa* hat auch die Fähigkeit, neue Resistenzen während Antibiotika-Behandlungen zu bilden (Narten et al., 2012). Das Bakterium kann Biofilme auf feuchten Oberflächen ausbilden, wie zum Beispiel Kathetern. Diese Biofilme können Ursache von komplizierten Harnwegsinfektionen (Tielen et al., 2013) sein.

Auch in Lungen von zystische Fibrose-Patienten, oder auch Mukoviszidose genannt (von lat. mucus „Schleim“ und viscidus „zäh, klebrig“), kann *P. aeruginosa* leben. Innerhalb der Lunge von zystische Fibrose-Patienten ist das Bronchialsekret zähflüssig und kann nicht mehr abtransportiert werden. Bei Patienten mit zystische Fibrose kommt es zu verstopften Atemwegen durch dickflüssigen Schleim, auch Mukus genannt (Lyczak et al., 2002). Innerhalb des Mukus kann eine komplexe Mikroflora entstehen und es kommt zu Entzündungen der Lunge. *P. aeruginosa* kann innerhalb der Lunge effizient Biofilme ausbilden. Da Teile des Mukus nicht mit Sauerstoff versorgt werden, ist es für *P. aeruginosa* essentiell, anaerob bzw. unter Bedingungen der reduzierten Sauerstoff-Konzentration überleben und wachsen zu können (Schobert and Jahn, 2010) (siehe Kapitel 6.2.1).

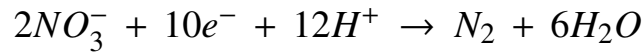
## 6.2. Denitrifikation in *P. aeruginosa*

Die Oxidation eines energiereichen Stoffes mit Hilfe der Reduktion von externen Elektronenakzeptors und die Speicherung der dabei freigesetzten Energie wird als Atmung bezeichnet. Die Denitrifikation ist eine anaerobe Atmung (Zumft, 1997), welche von *P. aeruginosa* und anderen fakultativ anaeroben Bakterien genutzt wird. Die Denitrifikation wird auch als Nitrataatmung bezeichnet. *P. aeruginosa* hat die Fähigkeit, ohne Sauerstoff zu überleben, kann aber auch Sauerstoff zur Atmung nutzen.

In dieser Atmungskette wird Nitrat mit der Nitrat- und der Nitritreduktase (NAR und NIR), der Stickstoffmonoxid-Reduktase (NOR) und der Distickstoffmonoxid-Reduktase (NOS)



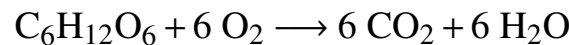
reduziert (Zumft, 1997). Die gesamte Denitrifikation



erzeugt molekularer Stickstoff ( $N_2$ ).

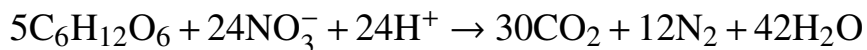
Die Denitrifikation ist eine Alternative zur oxidativen Atmung. Jedoch erzeugt die oxidative Atmung pro mol Glucose mehr Energie als die Denitrifikation (Strohm et al., 2007). Die theoretischen Netto-Energiebilanzen zeigt, dass die Oxidation immer energetisch günstiger ist.

### **Oxidation:**



$$\Delta G' = -2870 \text{ kJ pro mol Glucose}$$

### **Denitrifikation:**



$$\Delta G' = -2670 \text{ kJ pro mol Glucose}$$

Der Netto-Gewinn der Denitrifikation ist somit geringer als der Oxidation (Strohm et al., 2007). Die Denitrifikation ist jedoch energetisch deutlich besser als bei einer Fermentierung. Fakultativ anaerobe Mikroorganismen, z. B. *P. aeruginosa*, nutzen zunächst die Oxidation als Atmung. Ob Sauerstoff oder Nitrat als Elektronenakzeptor in der Atmungskette genutzt werden kann und welche Enzyme dem entsprechend produziert werden, wird über ein regulatorisches Netzwerk geregelt (siehe Kapitel 2.6.1). Ein regulatorisches Netzwerk in Bakterien kann die Expression von Genen steuern (siehe Kapitel 2.6).

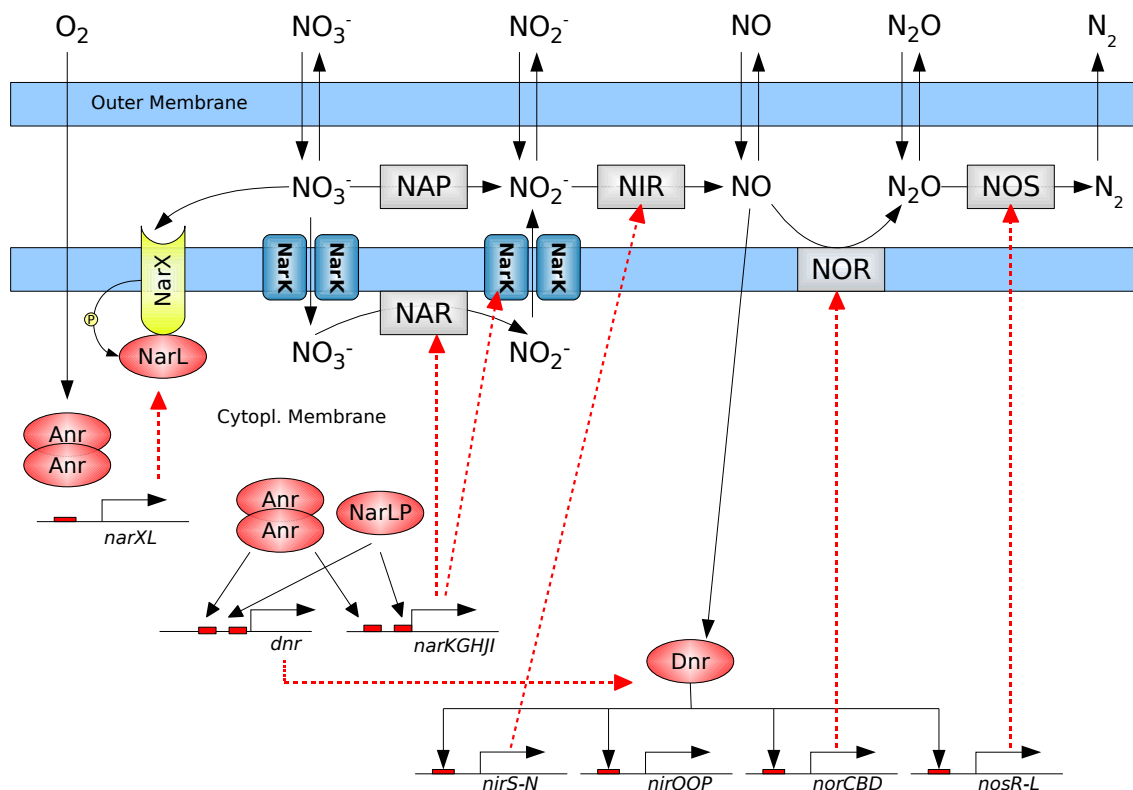


### 6.2.1. Regulation der Nitratatmung in *P. aeruginosa*

*P. aeruginosa* ist an sauerstoffarme Bedingungen angepasst. Mittels eines genregulatorischen Netzwerks steuert *P. aeruginosa* die Entscheidung über Sauerstoff- oder Nitratatmung (siehe Abbildung 6.1). Das Bakterium präferiert, wie die meisten fakultativ anaeroben Mikroorganismen, Sauerstoff als Energielieferant (Schobert and Tielen, 2010) (siehe Kapitel 6.2.1). Ist statt Sauerstoff nur Nitrat vorhanden, kann *P. aeruginosa* durch Denitrifikation Energie gewinnen.

In Abbildung 6.1 wird die Regulation der Nitratatmung in *P. aeruginosa* schematisch dargestellt. Das Protein Anr (*anaerobic regulation of arginine catabolism and nitrate reduction*) detektiert mittels eines Eisen-Schwefel-Cluster ( $[4Fe - 4S]^{2+}$ ) bei der Regulation zur Denitrifikation den Mangel an Sauerstoff (Trunk et al., 2010). Bei Sauerstoffabwesenheit bildet Anr ein Dimer aus (Yoon et al., 2007), siehe Abbildung 6.1 links. Das Eisen-Schwefel-Cluster wird bei Sauerstoff-Anwesenheit zerstört und Anr wird zu einem Monomer. Dimerisiertes Anr bindet an der sogenannten *Anr-Box* innerhalb des NarXL-Operons als Aktivator.

Das Zwei-Komponenten-System NarXL besteht aus der Sensorkinase NarX und dem Responseregulator NarL. Die Sensorkinase NarX ist in die innere Membran eingelagert (Abbildung 6.1 unterer blauer Balken, links), erkennt Nitrat und aktiviert NarL durch Phosphorylierung zu NarLP. Durch NarLP und dem dimerisiertem Anr kann die Expression des *dnr*-Gens und der *nar*-Gene aktiviert werden.



**Abbildung 6.1.: Regulatorisches Netzwerk der Denitrifikation von *P. aeruginosa*** (Münch, R., unpubliziert).

In Tabelle 6.1 finden sich Erläuterungen zu den beteiligten Komponenten. Zu sehen ist die äußere und innere Membran (hellblau) von *P. aeruginosa*. Hellrot sind Regulatoren gekennzeichnet. Rote Pfeile zeigen an, in welchem Kompartiment sich die Proteine befinden. Schwarze Pfeile stellen enzymatische Reaktionen oder Expressionen von Genen dar. Sauerstoff wird durch Anr detektiert, welches bei Abwesenheit von Sauerstoff durch das Eisen-Schwefel-Cluster dimerisieren kann und dann die Expression des Nitratsensorsystems NarXL aktiviert wird. NarX detektiert Nitrat und phosphoryliert bei Anwesenheit von Nitrat NarL zu NarLP. Mittels NarLP und dimerisierten Anr wird die Genkaskade der Nitratatmung aktiviert. Die Nitratreduktase NAR wird produziert. Mittels des Antiporters NarK wird Nitrat in und Nitrit aus der Zelle transportiert. Dimerisiertes Anr und NarLP aktivieren ebenfalls die Produktion von Dnr, welches als Aktivator für die Produktion der Untereinheiten von NIR (Nitritreduktase), NOR (Stickstoffmonoxid-Reduktase) und NOS (Distickstoffmonoxid-Reduktase) agiert. Mittels dieser drei Enzyme wird Nitrat schrittweise zu molekularem Stickstoff reduziert.

**Tabelle 6.1.: Elektronenakzeptoren, Regulationsenzyme und Gene der Nitratatmung** Übersicht der für die Nitratatmung in *P. aeruginosa* beteiligten elektronenakzeptierende Stoffe, Enzyme und Gene (Trunk et al., 2010, Yoon et al., 2007, Tielen et al., 2013).

Formel / Abkürzung	Name / Aufgabe
$O_2$	Sauerstoff, Elektronenakzeptor für aerobe Atmung
$NO_3^-$	Nitrat, Elektronenakzeptor für anaerobe Atmung
$NO_2^-$	Nitrit
$NO$	Distickstoffmonoxid
$N_2O$	molekularer Stickstoff
Anr	Dimerisiert bei Abwesenheit von $O_2$ . Regulator der <i>nar</i> -Gene und des <i>dnr</i> -Gen.
NarXL	Nitratsensorsystem
NarX	Membrangebundene Sensorkinase zur Detektion von $NO_3^-$ . Phosphoryliert NarL
NarL	Regulator für <i>dnr</i> -Gen und <i>nar</i> -Gene
NarLP	Phosphoryliertes NarL
NarK	Nitrat-Nitrit-Antiporter
Dnr	Regulator für <i>nir</i> -, <i>nor</i> - und <i>nos</i> -Genen
NAR	Nitratreduktase im Cytoplasma
NIR	Nitritreduktase
NAP	Dissimilatorische Nitratreduktase im Periplasma
NOS	Distickstoffmonoxid-Reduktase
NOR	Stickstoffmonoxid-Reduktase
<i>narXL</i>	Gen für Nitratsensorsystem
<i>dnr</i>	Gen für Regulator Dnr
<i>narK</i>	Gen für Antiporter
<i>narGHJI</i>	Gene für Membran gebundene Nitratreduktase
<i>nirS</i>	Gen für Cytochrom cd1 Nitritreduktase NirS
<i>nirN</i>	Möglich: Einbau vom Häm d1 in NirS
<i>nirQ</i>	Gen für regulatorisch wirkende Protein NirQ
<i>nirO</i>	Gen für Protein NirO
<i>nirP</i>	Gen für Protein NirP
<i>norBCD</i>	Gene für Membran gebundene Stickstoffmonoxid-Reduktase, <i>nor</i> Untereinheit C, B und D
<i>nosR</i>	Gen für regulatorisch wirkende Protein <i>NosR</i>
<i>nosL</i>	Gen für Protein <i>NosL</i>

Die *nar*-Gene kodieren die Nitratreduktase. Dnr als Regulator aktiviert die Expression der Gene für *nir*-, *nor*- und *nos*-Untereinheiten, welche die Nitratreduktase, die Stickstoffmonoxid-Reduktase und die Distickstoffmonoxid-Reduktase produzieren. Die Reduktasen bestehen ebenso wie die Nitratreduktase Nar aus mehreren Untereinheiten (siehe Tabellen 6.1).

Das Besondere an der Denitrifikationsregulation von *P. aeruginosa* ist die dauerhafte Induzierung des *narXL*-Operon. Die Expression findet auch in einer aeroben Umgebung statt, unabhängig vom Bedarf von NarXL. Mit bis zu 50 % von der Induktion unter anaeroben Bedingungen (Schreiber et al., 2007) wird das *narXL*-Operon unter aeroben Bedingungen induziert. Die Nitratreduktase wird nur exprimiert, wenn auch Anr als Dimer vorhanden ist. Daher wird trotz des exprimierten NarXL die Nitratreduktase in aerober Umgebung nicht produziert.

## 6.3. Feedforward-Schleife

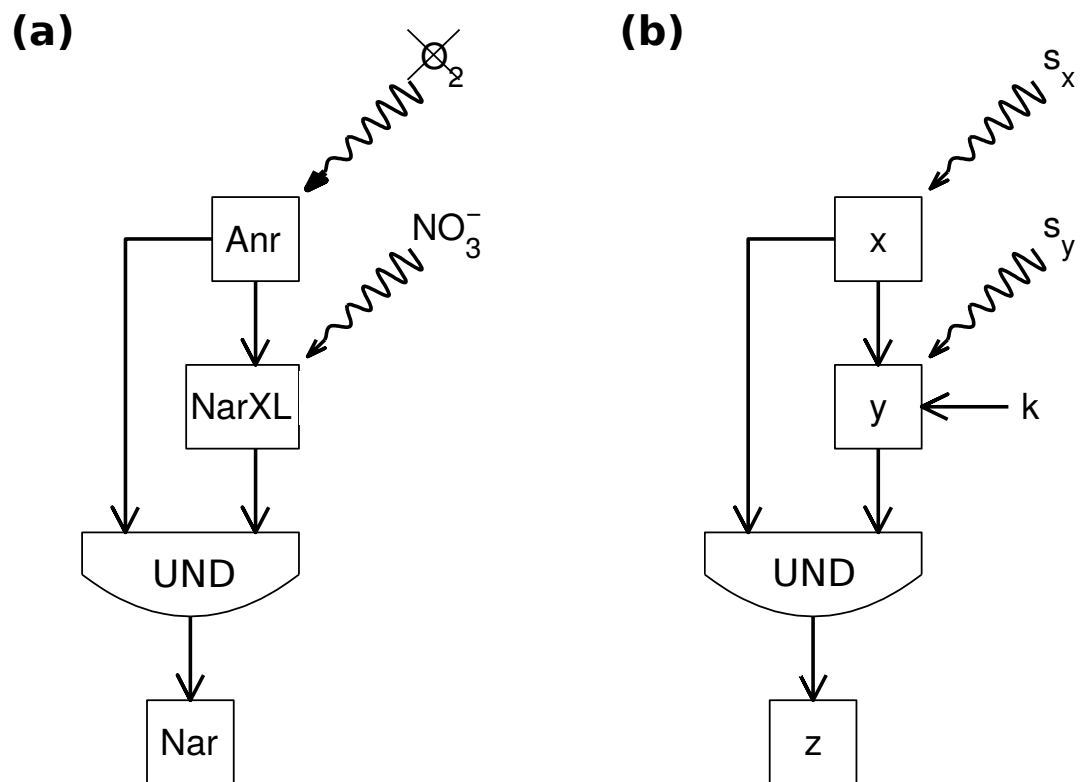
Eine FFL ist ein Netzwerkmotiv (Mangan and Alon, 2003) und besteht aus drei Zustandsvariablen ( $x(t)$ ,  $y(t)$ ,  $z(t)$ )<sup>1</sup>. Dabei wirken zwei der Zustandsvariablen als Regulatoren ( $x$  und  $y$ ) und eine Zustandsvariable symbolisiert die Zielvariable  $z$ , welche reguliert wird. Regulationen können aktivierend oder reprimierend wirken, es existieren durch binäre Kombinationen acht verschiedene FFL-Motive (Alon, 2007). Diese unterscheiden sich in ihrem dynamischen Verhalten und ihrer Robustheit (Macia et al., 2009, Widder et al., 2012).

Das FFL-Regulationsmotive findet man in der Nitratatmungsregulation von *P. aeruginosa* (Kapitel 6.2.1). Die Regulierung durch Anr und das Zwei-Komponenten-System NarXL sowie die gemeinsame darauffolgenden Akti-

---

<sup>1</sup> Variablen sind innerhalb dieser Arbeit bei nicht weiterer Kennzeichnung immer von der Zeit  $t$  abhängig.

vierung der Denitrifikations-Gene ist eine Typ-1 kohärente FFL (siehe Abbildung 6.2 (a)).



**Abbildung 6.2.: Typ-1 kohärente Feedforward-Schleife**

In (a) wird ein Teil von der Regulation der Nitratatmung in *P. aeruginosa* mittels Anr und NarXL vereinfacht dargestellt. Anr detektiert Sauerstoff  $O_2$ . Anr aktiviert gemeinsam mit Nitrat  $NO_3^-$  das NarXL-System. Das NarXL-System und Anr aktivieren gemeinsam die folgende Denitrifikation (als Nar gekennzeichnet). Diese Schleife wird Typ-1 kohärente Feedforward-Schleife genannt und lässt sich verallgemeinert in (b) darstellen. Hier werden die beiden Regulatoren und die darauffolgende Denitrifikation als drei Zustandsvariablen dargestellt. Anr wird als  $x$ , das NarXL-System als  $y$  und die Denitrifikation als  $z$  bezeichnet. Der erste Regulator  $x$  wird durch das Signal  $s_x$  aktiviert. Dieser reguliert dann die Produktion des zweiten Regulator  $y$  welcher durch das zweite Signal  $s_y$  aktiviert. Zusammen regulieren diese mit einer *UND*-Verknüpfung die Zielvariable  $z$ .

Innerhalb dieser Arbeit wird dieser Teil der Nitratatmungsregulation als verallgemeinerte Version betrachtet (siehe Abbildung 6.2 (b)). Der erste Regu-

lator  $x$  symbolisiert Anr, der zweite Regulator  $y$  das NarXL-System. Die Signale  $s_x$  und  $s_y$  stellen die Abwesenheit von Sauerstoff, bzw. die Anwesenheit von Nitrat dar. Die Proteine der darauffolgenden Denitrifikation wird in der verallgemeinerten Version mit  $z$  bezeichnet.

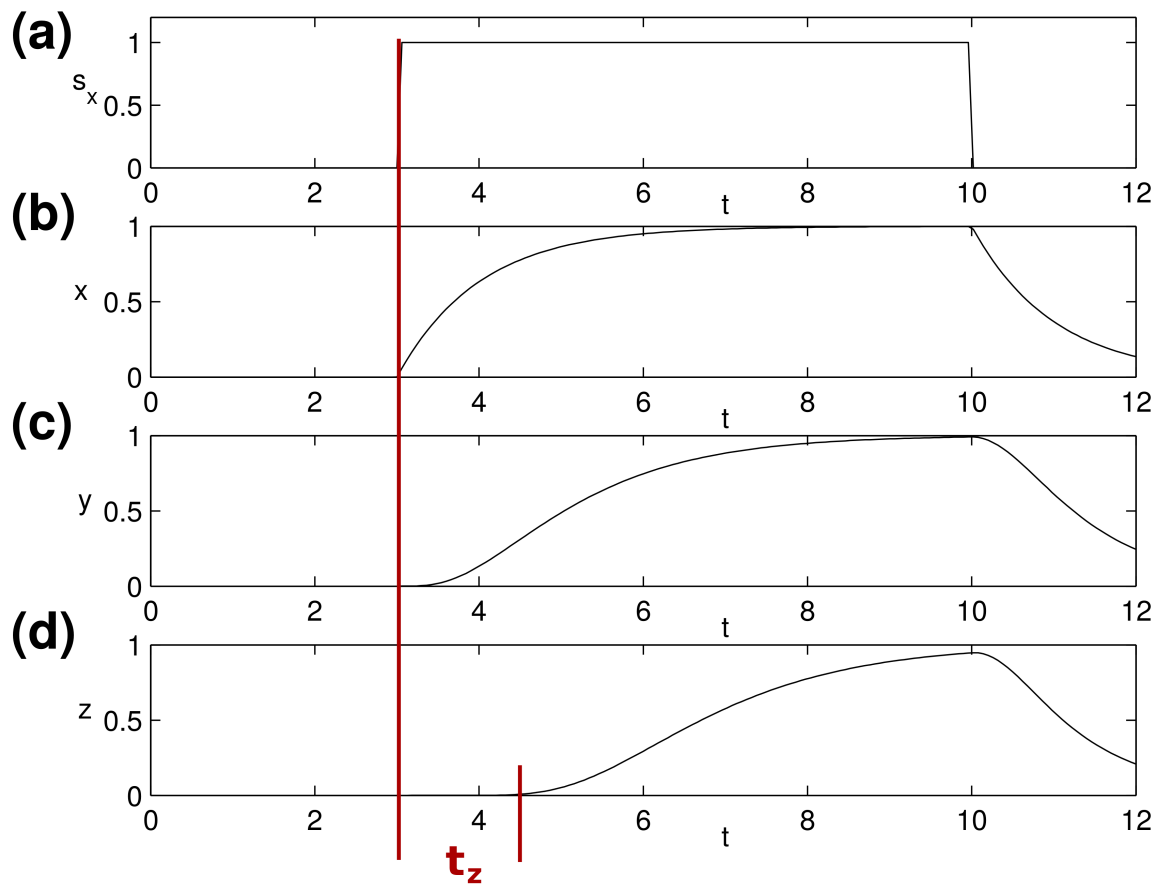
Bei der Typ-1 kohärente Feedforward-Schleife wird der erste Regulator  $x$  durch ein Signal ( $s_x(t)$ ) aktiviert und reguliert selbst den zweiten Regulator  $y$ , der zusätzlich durch ein zweites Signal  $s_y(t)$  aktiviert wird. Die Zielvariable wird produziert, wenn beide Regulatoren  $x$  und  $y$  zusammen vorhanden sind ( $UND$ ).

Nur bei Vorhandensein der beiden Signale  $s_x$  und  $s_y$  kommt es zu einem Wachstum der Zielvariable. Bei der Typ-1 kohärente FFL sind die regulatorischen Interaktionen daher nur Aktivierungen.

### 6.3.1. Filterung

Die Filterung von kurzen Signalen wie z. B. bei stark fluktuierenden Umweltbedingungen ist eine Möglichkeit für Bakterien, unnützen Energieaufwand und somit Energiekosten, zu verhindern. Die Typ-1 kohärente Feedforward-Schleife kann als Filter fungieren (Mangan et al., 2003).

Der erste Regulator  $x$  wird bei Anwesenheit des Signals  $s_x$  produziert. Die Filterung funktioniert zum einem über die Produktion des zweiten Regulators  $y$ , die erst nach erfolgreicher Produktion des ersten Regulators  $x$  effektiv gestartet wird. Zusätzlich hängt die Produktion der Zielvariable  $z$  von der Konzentration von beiden Regulatoren ab (Mangan and Alon, 2003, Kalir et al., 2005, Wall et al., 2005).



**Abbildung 6.3.: Filterung von Pulsen durch eine FFL**

Gezeigt werden die Zustandsvariablen und das Signal  $s_x$  einer generellen Typ-1 kohärenten FFL. Der erste Regulator  $x$  (b) wird bei Anwesenheit des Signals  $s_x$  (a) direkt produziert. Durch die Anwesenheit der ersten Regulator  $x$  wächst auch der zweite Regulator  $y$  (c). Die Produktion von  $z$  beginnt sofort (d), ist jedoch am Anfang sehr gering. Erst bei  $t_z$ , also wenn sowohl  $x$  als auch  $y$  groß genug sind, ist die Produktion von  $z$  nicht mehr vernachlässigbar.

Dadurch wird die Zielvariable  $z$  direkt nach Signalstart so gering produziert, dass dieses ignoriert werden kann (siehe Abbildung 6.3). Erst ab der Zeit  $t_z$  wird die Zielvariable stärker produziert. Innerhalb von  $t_z$  wird starke Produktion von  $z$  gefiltert.

### 6.3.2. Basale Expression des zweiten Regulators

Die Nitratatmungsregulation in *P. aeruginosa* weist eine spezialisierte FFL auf (siehe Kapitel 6.2.1). Der zweite Regulator NarXL ( $y$ ) in der FFL wird dauerhaft basal exprimiert (Schreiber et al., 2007). Die basale Expression von NarXL wird  $k$  genannt und liegt zwischen keiner und voller Produktion,  $k \in [0, 1]$ . Diese Regulation wird mit  $FFL_k$  bezeichnet im Gegensatz zur FFL ohne basaler Expression des zweiten Regulators, die  $FFL_0$  genannt wird. Eine  $FFL_k$ -Regulation wurde bereits in *E. coli* in der Arabinose-Regulation mittels AraC gezeigt (Mangan et al., 2003). Für *P. aeruginosa* ist es notwendig, sich möglichst gut an die Umgebung anzupassen, daher möglichst gut auf einen Wechsel von aerober zu anaerober Umgebung vorbereitet zu sein.

Die basale Expression der zweiten Zustandsvariable  $y$  kostet das Bakterium Energie. In aeroben Umgebungen bringt die basale Expression keinen Nutzen, da die Sauerstoffatmung verwendet wird, also lohnt sich dieser Aufwand ohne Signal nicht. Die basale Expression verringert jedoch die Zeit nach einem Signalwechsel, bis die Zielgene  $z$  vollständig produziert werden. Das Bakterium ist an sich schnell veränderbaren Umweltbedingungen angepasst und daher gut vorbereitet, wenn ein Signalwechsel stattfindet. Kosten und Nutzen der FFL mit basaler Expression hängen somit direkt von dem Wechsel der Umgebung ab.

Diese Arbeit beschäftigt sich damit, den evolutionären Vorteil einer basalen Expression der zweiten Zustandsvariable zu erläutern. Evolutionär vorteilhaft ist ein System, wenn das Verhältnis von Nutzen zu Kosten maximal und der Nutzen größer als die Kosten ist. Das Verhältnis von Nutzen zu Kosten soll abhängig von der basalen Expression  $k$  optimiert werden. Eine basale Expression ist nur dann vorteilhaft, wenn ihr Optimum zwischen voller



und keiner Produktion, also  $k \in ]0, 1[$ , liegt, daher ein strenges Maximum. Das Ziel dieser Arbeit ist es, ein möglichst einfaches Modell zu erstellen, welches den evolutionären Vorteil darstellen kann.

## **7. Modellierung der Feedforward-Schleife mit basaler Expression**

Es wird ein Differentialgleichungsmodell für das Problem des FFL-Systems mit basaler Expression des zweiten Regulators  $y$  aufgestellt (Roselius et al., 2014). Eine dazugehörige Nutzenbilanz-Funktion, welche das Verhältnis von Kosten und Nutzen zeigt, wird aufgestellt.

Gesucht wird eine Nutzenbilanz mit streng optimaler Lösung der basalen Expression. Hier wird der Einfachheit und Übersichtlichkeit halber das biologische System so einfach wie möglich gehalten. Variablen oder Parameter, die kaum einen Einfluss auf das System haben, werden vernachlässigt.

Die FFL kann durch ein Differentialgleichungssystem mathematisch dargestellt werden (siehe Kapitel 2.7). Es werden der Einfachheit halber nur die Proteine der Regulation betrachtet. Das Signal stellt den Sauerstoffgehalt der Umwelt stark vereinfacht dar.

## 7.1. Signal $s_x$

Das Signal  $s_x$  symbolisiert Sauerstoffabwesenheit. Daher ist  $s_x = 1$ , wenn kein Sauerstoff vorhanden ist und  $s_x = 0$ , wenn Sauerstoff vorhanden ist.

Das Signal  $s_x$  wechselt nur zwischen Sauerstoffanwesenheit und -abwesenheit,  $s_x(t) \in \{0, 1\}$ . Das Signal  $s_x$  ist abhängig von der Zeit, daher

$$s_x(t) = \begin{cases} 0 & \text{wenn Umgebung zum Zeitpunkt } t \text{ aerob,} \\ 1 & \text{wenn Umgebung zum Zeitpunkt } t \text{ anaerob ist.} \end{cases}$$

Das zweite Signal  $s_y$  symbolisiert die Anwesenheit von Nitrat. Nitrat wird als dauerhaft anwesend angenommen, daher ist das zweite Signal konstant ( $s_y = 1$ ).

### Definition 4.

Das Signal  $s_x$  ist eine stufenweise konstante Funktion. Die Zeitpunkte des Wechsels werden mit  $T_i$  bezeichnet, wobei  $T_{i+1} > T_i$  und  $T_0 = 0$  ist.

$$s_x(t) = \begin{cases} 0 & \text{wenn } t \in [T_{2i}, T_{2i+1}[ , i \in \mathbb{Z} \\ 1 & \text{wenn } t \in [T_{2i+1}, T_{2i+2}[ , i \in \mathbb{Z}. \end{cases} \quad (7.1)$$

## 7.2. Differentialgleichungsmodell

Die drei Zustandsvariablen in den drei Differentialgleichungen des FFL-Systems entsprechen der Konzentrationen beider Regulatoren  $x$  und  $y$  und der Zielvariablen  $z$ . Die Zustandsvariablen  $y$  und  $z$  hängen von der basalen Expression  $k$  ab und werden somit als  $y_k$  und als  $z_k$  bezeichnet. Der Aufbau (oder auch die Produktion) des ersten Regulators  $x$  erfolgt bei Anwesenheit des Signals  $s_x$ . Die Produktion des zweiten Regulators  $y_k$  wird

durch die basale Expression  $k$  und den ersten Regulator  $x$  gesteuert. Dieser Aufbau des zweiten Regulators wird durch die Funktion  $g_k(x)$  modelliert. Die Produktion der Zielvariablen  $z_k$  ist abhängig von beiden Regulatoren und wird durch die Funktion  $m(x, y_k)$  dargestellt. Um das System möglichst einfach zu verfassen, werden alle maximalen Zustände der Zustandsvariablen  $x$ ,  $y_k$  und  $z_k$  auf eins normiert. Daher wird festgelegt, dass  $\max(x) = \max(y_k) = \max(z_k) = 1$  ist. Die minimale Konzentration der Variablen ist Null,  $x, y_k, z_k \in [0, 1]$ <sup>1</sup>.

Die Produktionsfunktion des zweiten Regulators  $g_k(x)$  hängt von der ersten Zustandsvariable  $x$  und von der basalen Expression  $k$  ab. Die basale Expression  $k$  gibt an, wie stark die Produktion der zweiten Variable bei Abwesenheit des ersten Regulators  $x$  ist. Ist somit  $x = 0$ , so ist die basale Expression  $g_k(0) = k$ . Bei Maximum des ersten Regulators  $x = 1$  soll die Produktion des zweiten Regulators auch maximal mit  $g_k(1) = 1$  sein. Um so größer die basale Expression ist, desto stärker ist auch die Produktion des zweiten Regulators. Sind die Zustandsvariablen  $x$  und  $y_k$  maximal, so ist auch die Produktion  $m(x, y)$  von  $z_k$  maximal,  $m(1, 1) = 1$ .

Um so größer die Konzentration der ersten Zustandsvariable  $x$  ist, um so stärker wird die zweite Zustandsvariable  $y_k$  exprimiert. Die Produktionsfunktion  $g_k(x)$  ist eine stetige, monoton wachsende Funktion. Je größer die Konzentration  $x$  oder  $y_k$  sind, desto stärker wird  $z_k$  produziert. Somit ist die Produktionsfunktion  $m(x, y)$  ebenfalls eine stetige, monoton wachsende Funktion in beiden Komponenten.

### Definition 5.

Für die Produktionsfunktionen des zweiten Regulators  $y_k$  und der Zielvara-

---

<sup>1</sup>Ableitungen einer Funktion beziehen sich außer bei genaueren Angaben innerhalb dieser Arbeit immer nach der Zeit.

ble  $z_k$  gilt

$$\begin{aligned} g_k(0) &= k, \quad \forall k \in [0, 1], \\ g_k(1) &= 1, \quad \forall k \in [0, 1], \\ m(1, 1) &= 1. \end{aligned} \tag{7.2}$$

Die Funktionen  $g_k(x)$  und  $m(x, y_k)$  sind stetig monoton wachsende Funktionen

$$\begin{aligned} g_k(x_1) &> g_k(x_2), \quad \forall x_1 > x_2, \\ m(x_1, y_k) &> m(x_2, y_k), \quad \forall x_1 > x_2, \\ m(x, y_1) &> m(x, y_2), \quad \forall y_1 > y_2, \end{aligned} \tag{7.3}$$

in jeder Komponente. Weiterhin gilt, dass

$$g_k(x) > g_l(x) \geq 0 \text{ für } k > l \text{ und } \forall x \in [0, 1] \tag{7.4}$$

ist.

Der Abbau der Zustandsvariablen  $(x, y_k, z_k)$  wird durch deren eigene Konzentration gesteuert. Angenommen wird, dass die Abbauraten bei den Regulatoren und der Zielvariable gleich sind, um das Modell zu vereinfachen.

**Definition 6.**

Die FFL mit basaler Konzentration des zweiten Regulators wird mit dem Differentialgleichungssystem

$$\begin{aligned} \dot{x} &= s_x - x, \\ \dot{y}_k &= g_k(x) - y_k, \\ \dot{z}_k &= m(x, y_k) - z_k. \end{aligned} \tag{7.5}$$

dargestellt.

### 7.3. Nutzenbilanz-Funktion $C(k)$

Die FFL soll auf den Nutzen für das Bakterium hin untersucht werden. Es wird eine Nutzenbilanz-Funktion aufgestellt, welche die Kosten und den Nutzen zusammenfasst.

Nutzen bringt die Zielvariable  $z_k$  bei Anwesenheit des Signals  $s_x$ , in dem Beispiel der Nitratreduktase-Regulation von *P. aeruginosa* bringt die Nitratreduktase Nutzen bei Abwesenheit von Sauerstoff (Schobert and Tielen, 2010). Die Produktion von  $y_k$  und von  $z_k$  kostet Energie. Die Kosten der Dimerisierung von Anr, somit  $x$ , sind so gering, dass diese in der folgenden Betrachtung vernachlässigt werden können.

**Definition 7.**

Die Nutzenbilanz-Funktion  $C(k)$  wird definiert durch

$$C(k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt. \quad (7.6)$$

mit den Kostenfaktoren  $\eta_y$  und  $\eta_z$ .

Die Nutzenbilanz-Funktion ist der über die Zeit berechnete Mittelwert der Nutzen minus der Kosten. Sie hängt direkt von dem Signal  $s_x$  ab. Der Nutzen geht nur in die Funktion ein, wenn das Signal  $s_x$  vorhanden ist, die Produktionsfunktionen  $g_k(x)$  und  $m(x, y)$  hängen von den Zustandvariablen ab, dessen Produktion vom Signal abhängt. Die Nutzenbilanz hängt somit ebenfalls von dem Wechsel der Umgebung ab.

Die Nutzenbilanz-Funktion  $C(k)$  aus Definition 7 existiert, wenn

$$\bar{s} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s_x dt \in \mathbb{R} \quad (7.7)$$

existiert. Existiert  $\bar{s}$ , so existieren ebenfalls die Zustandsvariablen  $x$ ,  $y_k$  und  $z_k$  in Gleichung (7) und somit existiert auch die Gleichung (7).

Ist der Wechsel einer Umgebung periodisch, so ist das Signal mit einer periodischen Sequenz vorhanden.

**Satz 4.** *Ist das Signal mit einer  $T$ -periodischen Sequenz  $(T_i)_j$  vorhanden und ist der Unterschied  $(T_{i+1} - T_i \leq \text{const})$  beschränkt, so resultiert dies in  $T$ -periodischen Lösungen von  $x$ ,  $y_k$  und  $z_k$  und liefert damit die Nutzenbilanz-Funktion*

$$C(k) = \frac{1}{T} \int_0^T s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt. \quad (7.8)$$

**Beweis.** Der Limes der Nutzenbilanz-Funktion kann für Zeitintervalle dargestellt werden

$$C(k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt,$$

$$C(k) = \lim_{i \rightarrow \infty} \frac{1}{i} \sum_{j=1}^{(i-1)} \frac{1}{T_{j+1} - T_j} \int_{T_j}^{T_{j+1}} s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt.$$

Für eine gegebene Periode  $T$  gilt, dass  $T_{j+1} - T_j = T$  ist. Wenn das Signal periodisch ist, so kommt es ebenfalls zu periodischen Lösungen der Zustandsvariablen. Die Zeit kann in Intervalle unterteilt werden, die jeweils  $T$  groß

sind. Damit ist

$$C(k) = \lim_{i \rightarrow \infty} \frac{1}{i} \sum_{j=1}^i \frac{1}{T} \int_0^T s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt.$$

$$C(k) = \lim_{i \rightarrow \infty} \frac{1}{i} i \left( \frac{1}{T} \int_0^T s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt. \right)$$

$$C(k) = \lim_{i \rightarrow \infty} \frac{1}{T} \int_0^T s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt.$$

Das Integral ist unabhängig von  $i$  und es gilt für  $T$ -periodische Signale

$$C(k) = \frac{1}{T} \int_0^T s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt.$$

□

Die Nutzenbilanz zeigt das Verhältnis von Nutzen und Kosten über den Zeitraum  $T$ . Soll dieses Verhältnis nur zu einem Zeitpunkt dargestellt werden, wird die momentane Nutzenbilanz benötigt.

**Definition 8.** Die *momentane* Nutzenbilanz

$$c(k) = s_x z_k - \tilde{c}_k = s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) \quad (7.9)$$

wird durch den Integranden aus Gleichung (7.8) dargestellt.  $\tilde{c}_k$  bezeichnet die momentanen Kosten.

Mittels der Nutzenbilanz-Funktion kann berechnet werden, ob der Nutzen, der aus dem Modell gezogen werden kann, größer ist als die Kosten.



**Definition 9.** Innerhalb dieser Arbeit ist ein System *evolutionär vorteilhaft*, wenn die Nutzenbilanz  $C(k)$  nicht sinkt, also die momentane Nutzenbilanz nur positiv und wenn die Nutzenbilanz abhängig von der basalen Expression streng maximal  $\max(C(k))$  ist.

Ist die Nutzenbilanz nicht wachsend, so kann das Bakterium möglicherweise ebenfalls nicht wachsen. Es könnte zu einem Sterben des Bakteriums kommen. Ist die Nutzenbilanz nicht maximal, so kann es andere Bakterien geben, die besser vorbereitet sind als diese .

## 7.4. Voraussetzungen

Um den Nutzen der basalen Expression  $k$  des zweiten Regulators  $y_k$  zu diskutieren und eine basale Expression  $k$  zu finden, bei der die Nutzenbilanz-Funktion optimal ist, muss ein Lemma formuliert werden:

### Lemma 1.

*Eine Differentialgleichung mit dem Aufbau*

$$\dot{u} = f - u \quad (7.10)$$

*und einer Anregung  $f$  hat die Lösung*

$$u = e^{-t}u(0) + \int_0^t e^{\tau-t} f(\tau) d\tau. \quad (7.11)$$

*Ist die Funktion  $f$  eine  $T$ -periodische Funktion, so kommt es zu*

$$L[f] = u = \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} f(\tau) d\tau + \int_0^t e^{\tau-t} f(\tau) d\tau. \quad (7.12)$$

**Beweis.** Man kann direkt aus Gleichung (7.10) die Funktion (7.12) berechnen. Die Lösung von Gleichung (7.10) ist

$$u(t) = \int_0^t f(\tau) e^{\tau-t} + C e^{-t}.$$

Fügt man  $u(0)$  als Startwert ein, so ist

$$u(t) = e^{-t} u(0) + \int_0^t e^{\tau-t} f(\tau) d\tau.$$

Gehen wir von einer  $T$ -periodischen Lösung aus, ist

$$u(0) = u(T) = e^{-T} u(0) + \int_0^T e^{\tau-T} f(\tau) d\tau.$$

Man kann

$$u(0) = \frac{1}{1 + e^{-T}} \cdot \int_0^T e^{\tau-T} f(\tau) d\tau$$

berechnen und in die Gleichung (7.11) einfügen,

$$u = e^{-t} u(0) + \int_0^t e^{\tau-t} f(\tau) d\tau = e^{-t} \frac{1}{1 + e^{-T}} \cdot \int_0^T e^{\tau-T} f(\tau) d\tau + \int_0^t e^{\tau-t} f(\tau) d\tau.$$

Definiert wird die Funktion dann als

$$L[f] = u = \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} f(\tau) d\tau + \int_0^t e^{\tau-t} f(\tau) d\tau.$$

□

Die Gleichung  $L[f] = u$  (Gleichung (7.12)) definiert einen linearen Operator  $L : \mathcal{L}_1([0, T]) \rightarrow \mathcal{L}_1([0, T])$ . Damit  $L$  ein linearer Operator ist, muss gelten, dass  $L$  homogen und additiv ist. Homogen ist  $L$ , wenn gilt, dass

$$L[\lambda x(t)] = \lambda L[x(t)]$$

ist. Man kann dies zeigen durch

$$\begin{aligned} L[\lambda \cdot x(t)] &= \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} (\lambda \cdot x(\tau)) d\tau + \int_0^t e^{\tau-t} (\lambda \cdot x(\tau)) d\tau \\ L[\lambda \cdot x(t)] &= \lambda \cdot \left( \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} (x(\tau)) d\tau + \int_0^t e^{\tau-t} (x(\tau)) d\tau \right) = \lambda L[x(t)]. \end{aligned}$$

Additiv ist  $L$ , wenn gilt, dass

$$L[x(t) + y(t)] = L[x(t)] + L[y(t)]$$

Daher ist

$$L[x(t) + y(t)] = \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} (x(\tau) + y(\tau)) d\tau + \int_0^t e^{\tau-t} (x(\tau) + y(\tau)) d\tau$$

und

$$\begin{aligned} L[x(t) + y(t)] &= \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} x(\tau) d\tau + \int_0^t e^{\tau-t} x(\tau) d\tau + \\ &\quad \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} y(\tau) d\tau + \int_0^t e^{\tau-t} y(\tau) d\tau = L[x(t)] + L[y(t)]. \end{aligned}$$

**Satz 5.**

Mit Lemma 1 lassen sich die Zustandsvariablen  $x(t), y_k(t)$  und  $z_k(t)$  aus den Gleichungen (7.5) durch die Funktion  $L$  darstellen.

$$x = L[s_x]$$

$$y_k = L[g_k(x)]$$

$$z_k = L[m(x, y_k)]$$

**Lemma 2.** Die Lösungen  $(x, y_k, z_k)$  und  $(x, y_l, z_l)$  zu dem System der Gleichungen (7.5) zeigen, dass  $y_k(t) \geq y_l(t)$  und  $z_k(t) \geq z_l(t)$  für alle  $t > 0$  ist unter der Voraussetzung, dass  $k \geq l \geq 0$ ,  $y_k(0) \geq y_l(0)$  und  $z_k(0) \geq z_l(0)$  sind.

**Beweis.** Durch Gleichung (7.11) aus Lemma 1 wird die Monotonie von  $g_k(x)$  in  $k$  gegeben und dieses impliziert die Monotonie der Lösung  $y_k(t)$  in  $k$ . Zusätzlich hängt  $y_k$  monoton steigend von dem Anfangswert  $y_k(0)$  ab. Analog hängt die Monotonie von  $z_k$  von der Monotonie  $m(x, y_k)$  in  $y_k$  ab, welche ebenfalls mit dem Anfangswert ( $z_k(0)$  und  $k$ ) steigt.  $\square$

**Satz 6.** Bei einem periodischen Signal  $s_x$  zeigt sich Monotonie von  $y_k = L[g_k(x)]$  und  $z_k = L[m(x, y_k)]$  in  $k$ .

**Beweis.** Es gilt nach (7.4), dass  $g_k(x) > g_l(x) \geq 0$  für  $k > l$  und  $\forall x \in [0, 1]$ . Daher ist  $g_k$  monoton wachsend in  $k$ . Gezeigt werden kann die Funktion:

$$L[g_k] = \frac{e^{-t}}{1 - e^{-T}} \int_0^T e^{\tau-T} g_k(\tau) d\tau + \int_0^t e^{\tau-t} g_k(\tau) d\tau.$$

Nur die Variable  $g_k$  ist abhängig von  $k$ , welche nur positiv in die Gleichung eingeht. Somit ist  $L[g_x]$  monoton wachsend in  $k$ . Es wird genauso

für  $z_k = L[m(x, y_k)]$  vorgegangen.  $\square$

Daher hat ein Organismus mit einer basalen Produktion von  $y_k$  ( $k > 0$ ) eine höhere Konzentration des Regulators  $y_k$  als bei  $k = 0$  und somit eine höhere Expression von  $z_k$  direkt nach einer Umschaltung für jedes periodische Gleichgewicht und jedes Systemverhalten mit gleichen Anfangsbedingungen.

## 8. Ergebnisse und Diskussion

Dieses Kapitel soll ein mathematisches Modell für das FFL-System mit basaler Expression des zweiten Regulators  $y_k$  liefern. Gesucht wird ein Modell, welches die evolutionären Vorteile der basalen Expression des zweiten Regulators erläutern kann. Das FFL-Modell wird schrittweise derart erweitert, dass ein möglichst einfaches Modell entsteht. Alle hier durch Erweiterungen entstehenden Modelle gehören zu einer Modellfamilie, welche gleichen Voraussetzungen aus Kapitel 7 erfüllen. Dazu müssen die Bedingungen des evolutionären Vorteils aus Definition 9 gezeigt werden.

### 8.1. Lineares Modell

Das Modell hat, wie in Kapitel 7 beschrieben, den Grundaufbau aus den Gleichungen (7.5) mit den Produktionsfunktionen  $g_k(x)$  und  $m(x, y_k)$ . Zunächst werden die Produktionsfunktionen von  $y_k$  und  $z_k$  als lineare Funktionen in jeder Komponente definiert

$$g_k(x) = k + (1 - k)x, \tag{8.1}$$

$$m(x, y_k) = x \cdot y_k. \tag{8.2}$$

mit den Voraussetzungen aus Definition 5 in Kapitel 7.2.

Das Modell aus den Gleichungen (7.5) mit den Produktionsfunktionen aus Gleichung (8.1) und (8.2) wird als lineares Modell bezeichnet. Das Signal  $s_x$  wurde in Definition 7.1 festgelegt, wobei nun periodische Zeitintervalle genutzt werden. Es gilt, dass  $T_{i+2} - T_i = T_2 \forall i \in \mathbb{Z}$  mit  $T$  einer endlichen Periodenlänge.

### 8.1.1. Nutzenbilanz bei Signalwechsel

Um die Nutzenbilanz bei Signalwechsel zu definieren, wird das Landau-Symbol  $O$  verwendet (siehe Einschub 1).

**Einschub 1.** Die asymptotische obere Schranke des Wachstums kann angegeben werden mittels des Landau-Symbol  $O$ . Für gegebene die Funktionen  $f$  und  $g$  gilt für  $f \in O(g)$ , dass bei  $x \rightarrow \infty$

$$\exists C > 0 \exists x_0 > 0 \forall x > x_0 : \left| \frac{f(x)}{g(x)} \right| \leq C$$

und für  $x \rightarrow a < \infty$  gilt

$$\exists C > 0 \exists \varepsilon > 0 \forall x \in \{x : d(x, a) < \varepsilon\} : |f(x)| \leq C \cdot |g(x)|.$$

$f$  hat als asymptotische Wachstumsobergrenze  $g$ . Strebt  $|f(x)|/|g(x)|$  gegen 0, so ist entsprechend  $f(x) \in o(g(x))$ .

**Lemma 3.** *Das Differentialgleichungssystem aus Kapitel 8.1 mit den Anfangsbedingungen  $x(0) = 0$ ,  $y_k(0) = k$  und  $z_k(0) = 0$  und den linearen Produktionsfunktionen (Gleichung (8.1) und Gleichung (8.2)) wächst beim Auftreten des ersten Signales, also zum Zeitpunkt  $T_1$ , für die Zeitlängen  $\tau$ .*

$$x(T_1 + \tau) = O(\tau) \text{ und } y_k(T_1 + \tau) = k + O(\tau^2).$$

Die zusätzliche Annahme, dass  $m(x, y_k) = O(xy)$  ist, ergibt folgendes:

$$z_k(T_1 + \tau) = kO(\tau^2) + O(\tau^4).$$

**Beweis.** Das Lemma nutzt die linearen Produktionsfunktionen  $g_k(x)$  und  $m(x, y)$  aus Gleichung (8.1) und Gleichung (8.2) und kann mittels des Hartman-Grobman-Theorems bewiesen werden (Hartman, 1960).

Das Hartman-Grobman-Theorem, oder auch Linearisierungssatz, analysiert das lokale Verhalten von dynamischen Systemen. Ein dynamisches System verhält sich in der Umgebung eines hyperbolischen Fixpunktes qualitativ wie ein lineares System.

Die Anfangsbedingungen wurden angegeben durch  $x(0) = 0$ ,  $y_k(0) = k$  und  $z_k(0) = 0$ . Zum Zeitpunkt  $T_1$  tritt das Signal zum ersten Mal auf, daher gilt ebenfalls zum Zeitpunkt  $T_1$ , dass  $x(T_1) = 0$ ,  $y_k(T_1) = k$  und  $z_k(T_1) = 0$ . Mit Hilfe der Taylorentwicklung wird der Zeitpunkte  $t_0 = T_1 + \tau$  mit der Entwicklungsstelle  $t = T_1$  betrachtet,

$$x(t_0) = x(t)(t_0 - t) + O(t_0 - t),$$

$$x(T_1 + \tau) = x(T_1)(T_1 + \tau - T_1) + O(T_1 + \tau - T_1),$$

$$x(T_1 + \tau) = 0 + O(\tau),$$

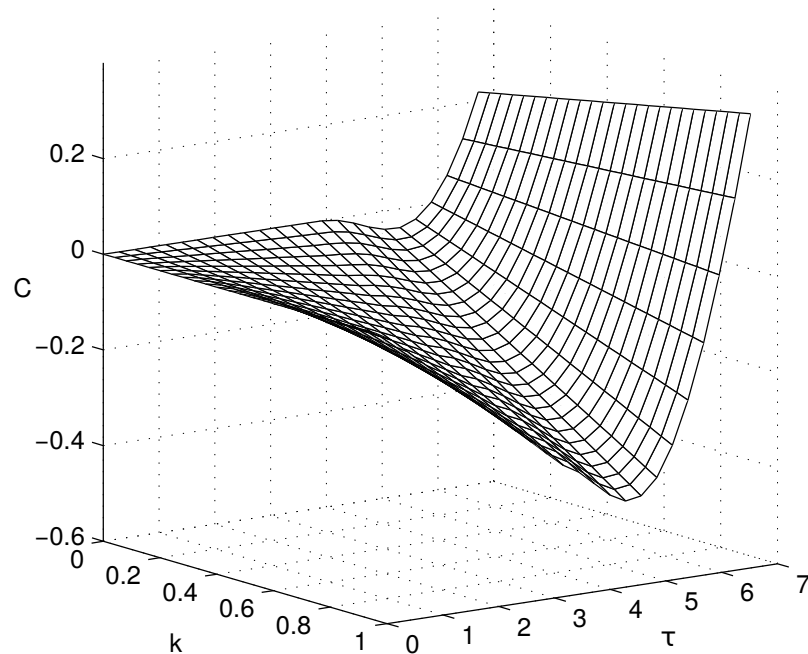
$$x(T_1 + \tau) = O(\tau).$$

Ebenso wird für die anderen Variablen vorgegangen.  $\square$

Lemma 3 zeigt, dass in jeder Situation direkt nach dem ersten Wechsel des Signals, also bei  $T_1 + \tau$ , die momentanen Kosten  $\tilde{c}_k = \eta_y g_k(x) + \eta_z m(x, y_k) = \eta_y k + O(\tau)$  wesentlich höher sind als der momentane Nutzen  $s_x(T_1 + \tau)z_k(T_1 + \tau)$ . Wechselt also die Umwelt und somit das Signal, kommt es zunächst zu



erhöhten momentanen Kosten. Das Bakterium benötigt also bei einem Umgebungswechsel zunächst Energie. Die benötigte Energie nach dem Signalwechsel wird in Abbildung 8.1 dargestellt.



**Abbildung 8.1.: Die Nutzenbilanz direkt nach dem ersten Wechsel des Signals innerhalb des linearen Modells**

Die Nutzenbilanz des linearen Modells  $C(k)$  dargestellt direkt nach dem Wechsel des Signals bei  $\tau = t - T_1$ . Die Kosten-Faktoren werden als  $\eta_y = \eta_z = 0.1$  festgelegt. Die Nutzenbilanz-Funktion ist negativ für kleine  $\tau$ . Bei größeren  $\tau$  wächst der momentane Nutzen stärker als die momentanen Kosten. Somit wächst auch  $C(k)$ .

Der Simulations-Start befindet sich bei  $T_1$  mit  $C(k) = 0$  zu dem Zeitpunkt  $T_1$ , mit der Zeit nach dem ersten Signalwechsel  $t = T_1 + \tau$  und variierenden basalen Expressionen  $k$ . Hier kann deutlich erkannt werden, dass die Nutzenbilanz innerhalb dieses Modells negativ sein kann. Das Modell kann somit die evolutionäre Vorteil innerhalb dieses Zeitbereiches nicht erklären. Von einem evolutionären Vorteil eines Modells kann dann ausgegangen werden, wenn die Nutzenbilanz sowohl maximal als auch wachsend

ist (siehe Definition 9). Die sinkende Nutzenbilanz innerhalb dieses Bereiches nach dem Wechsel zeigt, dass das Bakterium möglicherweise einen Wechsel der Umgebung nicht überleben könnte. Es wird im weiteren untersucht, ob das Modell die Optimalität einer basalen Expression  $k$  erläutern kann.

### 8.1.2. Maximum der basalen Expression

Das lineare Modell der FFL kann nicht den evolutionären Vorteil (siehe Definition 9) der basalen Expression erklären. Es existiert kein strenges Maximum von  $k \in ]0, 1[$  in der Nutzenbilanz-Funktion  $C(k)$  und die momentane Nutzenbilanz ist nicht nur positiv. Es wird gezeigt, dass  $C(k)$  linear in  $k$  ist. Ist dies der Fall, kann das Maximum der Nutzenbilanz-Funktion nur bei  $k = 0$  oder  $k = 1$  liegen.

**Satz 7.** *Das lineare Modell hat kein streng optimales  $k$  der Nutzenbilanz-Funktion  $C(k)$ , also kein optimales  $k$  mit  $k \in ]0, 1[$ .*

**Beweis.** Für den Beweis wird der Operator  $L$  aus der Gleichung (7.12) genutzt (siehe Lemma 1). Die Lösungen des Differentialgleichungsmodells aus Kapitel 8.1 können mittels dem Operator  $L$  und  $y_0 = L(x)$  berechnet werden,

$$x = L[s_x], \quad (8.3)$$

$$y_k = L[k + (1 - k)x], \quad (8.4)$$

$$z_k = L[xy_k] = L[x(k + (1 - k)y_0)] = ky_0 + (1 - k)L[y_0x]. \quad (8.5)$$

Damit kann die Nutzenbilanz-Funktion  $C(k)$  explizit angegeben werden:

$$C(k) = \frac{1}{T} \int_0^T s_x (ky_0 + (1-k)L[y_0x]) - \eta_y (k + (1-k)x) - \eta_z x (k + (1-k)y_0) dt.$$

Die Nutzenbilanz-Funktion wird derart umgestellt, dass die Linearität in  $k$  deutlich sichtbar wird.

$$\begin{aligned} & \frac{1}{T} \int_0^T s_x (ky_0 + (1-k)L[y_0x]) - \eta_y (k + (1-k)x) - \eta_z x (k + (1-k)y_0) dt \\ &= \frac{1}{T} \int_0^T s_x ky_0 + s_x L[y_0x] - s_x kL[y_0x] - \eta_y k - \eta_y x + \eta_y kx - \eta_z xk - \eta_z xy_0 + \eta_z ky_0 dt \\ &= \frac{1}{T} \int_0^T k \cdot (s_x y_0 - s_x L[y_0x] - \eta_y + \eta_y x - \eta_z x + \eta_z y_0) - \eta_y x + s_x L[y_0x] - \eta_z xy_0 dt \end{aligned}$$

Die Nutzenbilanz-Funktion hat mit linearen Produktionsfunktionen die Form

$$C(k) = \frac{1}{T} \int_0^T Q_1(t, s_x, y_0, \eta_y, \eta_z)k + Q_2(t, s_x, y_0, \eta_y, \eta_z) dt,$$

wobei

$$Q_1 = s_x y_0 - s_x L[y_0x] - \eta_y + \eta_y x - \eta_z x + \eta_z y_0$$

und

$$Q_2 = -\eta_y x + s_x L[y_0x] - \eta_z xy_0$$

Funktionen abhängig von den Parametern und Variablen, jedoch nicht von

der basalen Expressionsrate  $k$ , sind. Die Nutzenbilanz-Funktion ist linear abhängig von  $k$ . Somit kann für dieses Modell kein strenges Maximum in  $k$  existieren, sondern nur im Minimum oder im Maximum der basalen Expression, also  $k \in \{0, 1\}$ , in Abhängigkeit des Vorzeichens von  $Q_1$ . Das Vorzeichen von  $Q_1$  hängt von seinen Variablen, nämlich  $s_x, y_0, \eta_y$  und  $\eta_z$  ab.  $\square$

### 8.1.3. Simulation mit linearem Modell

Ein Beispiel für die Simulation mit dem linearen Modell bei periodischer Änderung des Signals wird in Abbildung 8.2 gezeigt. Die Abbildung verdeutlicht den Übergang zwischen Phasen mit und ohne Signale. Innerhalb der Simulation des linearen Modells werden periodische Signale angenommen mit einer Periodenlänge von  $T = 5$  und einem Umschaltzeitpunkt nach der Hälfte des Signals. Insgesamt werden vier Zyklen gezeigt:

$$s_x = \begin{cases} 0 & \text{wenn } t \in [5 \cdot (i - 1), 2.5 + 5(i - 1)[ \\ 1 & \text{wenn } t \in [2.5 + 5(i - 1), 5 \cdot i[ \end{cases} \quad i \in \{1, 2, 3, 4\}. \quad (8.6)$$

In der Abbildung 8.2 (a) wird der erste Regulator  $x$  dargestellt. In grau abgehoben wird die Anwesenheit des Signals gezeigt. Bei einem Stimulus wächst  $x$  exponentiell. Bei Abwesenheit des Signals wird  $x$  abgebaut.

Der zweite Regulator  $y_k$  wird basal exprimiert. In Abbildung 8.2 - (b) bis (d) werden sechs verschiedene basale Expressionen ( $k \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ ) dargestellt. Bei Anwesenheit des Signals wächst sowohl der erste als auch der zweite Regulator ( $x$  und  $y_k$ ). Der zweite Regulator wächst abhängig von der basalen Expression. Eine höhere basale Expression führt zu einer größeren Konzentration von  $y_k$ , wenn es noch nicht das Maximum erreicht hat.

Bei längeren Signalen tendiert der zweite Regulator, unabhängig von der basalen Expression, zu seinem Maximum ( $y_k = 1$ ).

Die Produktionsfunktion der Zielvariable  $z_k$  ist abhängig von der Konzentration von  $x$  und  $y_k$ .  $z_k$  wächst nur bei Anwesenheit von  $x$  und  $y_k$ , somit nur bei Anwesenheit des Signals  $s_x = 1$ . Durch eine höhere basale Expression ist es der Zielvariablen möglich, direkt nach dem Wechsel zur Signalanwesenheit (bei  $T_1 \geq 2.5 + 5(i - 1)$  für  $i \in \{1, 2, 3, 4\}$ ) schneller zu wachsen. Je höher die basale Expression  $k$  ist, desto schneller wächst ist die Konzentration von  $z_k$  nach einem Signalwechsel, gezeigt im Lemma 3.

Sind die Variablen noch nicht vollständig abgebaut und es kommt erneut zu einem Signal  $s_x = 1$ , so kann der erneute Aufbau der Variablen weniger Zeit dauern, als bei vollständig abgebauten Variablen. Solange die Variablen noch mit einer gewissen Konzentration vorhanden sind, *erinnert* sich das Bakterium sozusagen an die vorherige Umgebung.

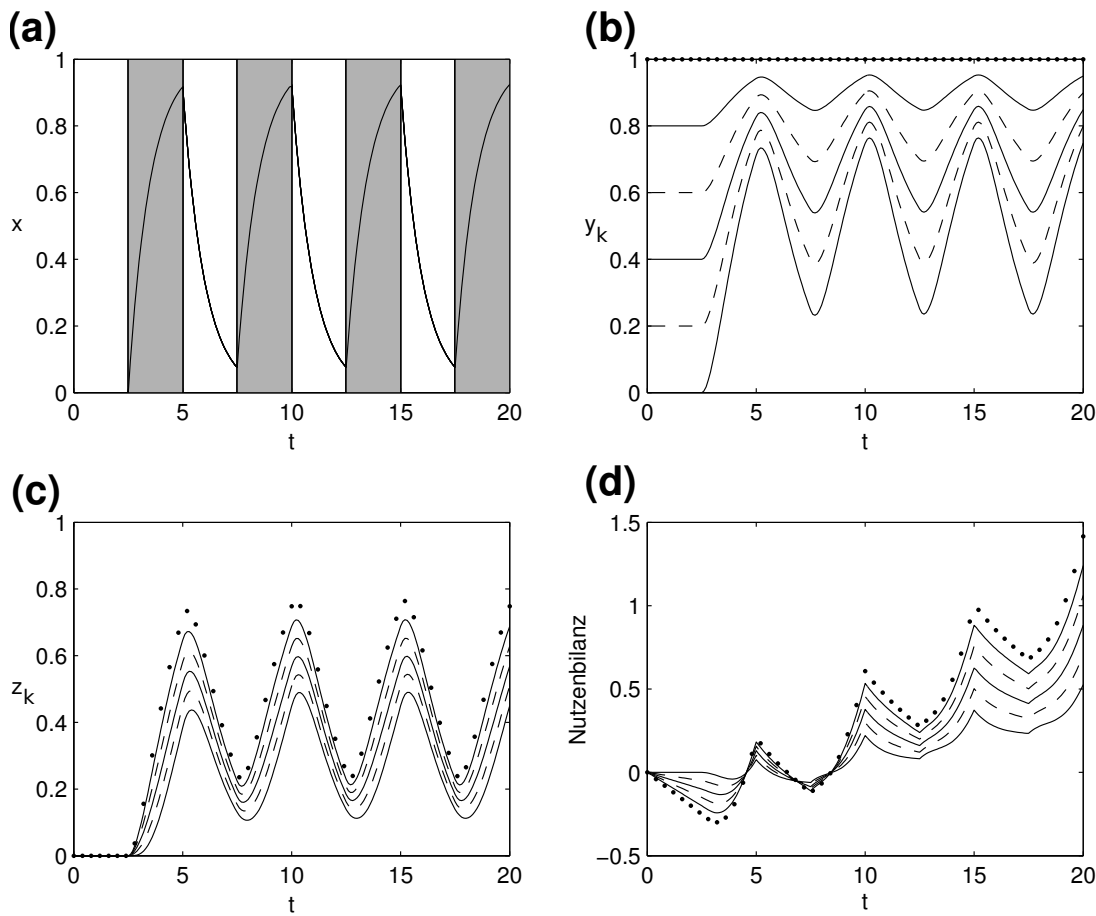
**Definition 10.** Das **Gedächtnis** wird hier die Zeit genannt, die gebraucht wird, bis die Variablen vollständig abgebaut sind. Je schneller der Abbau der Variablen ist, desto kürzer ist auch das Gedächtnis. Sind alle Variablen wieder vollständig abgebaut, ist kein Gedächtnis mehr vorhanden.

Die basale Expression hat direkten Einfluss auf das Gedächtnis. Um so größer die basale Expression ist, um so länger ist auch Gedächtnis. Bei kurzen Signalen hilft also der verzögerte Abbau, der auch durch die basale Expression entsteht, möglichst schnell wieder einen Nutzen zu bekommen. Bei längeren Signalen gleichen sich die Konzentrationen der Zustandvariablen an, so dass das Gedächtnis nicht mehr vorhanden ist. Das Bakterium *vergisst*.

Die Nutzenbilanz wird in Abbildung 8.2 (d) gezeigt. Kosten bei Signalabwesenheit sind größer als der Nutzen. Die Nutzenbilanz wächst nur mit Stimulus ( $s_x = 1$ ).

Liegt eine größere basale Expression vor, so sind ebenfalls die Kosten größer

$$\eta_y \cdot (k + (1 - k)x) - \eta_z \cdot (xy_k) > \eta_y \cdot (l + (1 - l)x) - \eta_z \cdot (xy_l) \quad \forall k > l.$$



**Abbildung 8.2.: Simulationsergebnis des linearen Modells.**

Die Phasen mit Signal werden durch einen grauen Hintergrund bei der Zustandsvariablen  $x$  dargestellt (a).  $x$  wächst in Anwesenheit des Signales  $s_x$ . In (b) wird der zweite Regulator  $y_k$  für die basale Expression  $k = 0, 0.2, 0.4, 0.6, 0.8$  (durchgängige, gestrichelte, durchgängige, gestrichelte Linie) und  $k = 1$  (gepunktete Linie) gezeigt. Die Darstellung der Zielvariablen  $z_k$  befinden sich in der Teilabbildung (c). Die Kosten-Faktoren der Nutzenbilanz  $C(k)$  werden  $\eta_y = \eta_z = 0.1$  gesetzt (d).

Liegt eine größere basale Expression vor, kann schneller aus dem Signal Nutzen gezogen werden als bei einer kleineren basalen Expression. Dies kommt durch den schnelleren Aufbau der Zielvariable bei erneutem Signal.

In den Phasen des Wechsels zwischen Ab- und Anwesenheit des Signals hat ein Modell mit einer größeren basalen Expression einen so großen Nutzen, dass es die Kosten wieder aufwiegt, und somit Modelle mit kleinerer basaler Expression in der Nutzenbilanz-Analyse 'überholt'. Erkennbar ist, dass immer eins von den Modellen mit fehlender basaler Expression ( $k = 0$ ) oder mit vollständiger basaler Expression ( $k = 1$ ) die maximale Nutzenbilanz-Funktion besitzt.

Das lineare Modell kann nicht den evolutionären Vorteil (siehe Definition 9) einer basalen Expression zwischen eins und null ( $k \in ]0, 1[$ ) erklären. Im folgendem wird das Modell erweitert, um einen evolutionären Vorteil durch ein Modell zeigen zu können.

## 8.2. Nichtlineares Modell

Um in dem Modell eine evolutionär stabile basale Expression des zweiten Regulators zu erläutern, werden Erweiterungen an dem Modell vorgenommen. Eine Möglichkeit der Erweiterung des Differentialgleichungsmodells der FFL (Gleichung (7.5)) ist die Veränderung der Produktionsfunktionen des zweiten Regulators  $y_k$  und der Zielvariablen  $z_k$ . Für die Produktionsfunktionen  $g_k(x)$  und  $m(x, y_k)$  wird nun eine Sättigungsfunktion mit Hill-Koeffizient angewendet. Dieses Modell wird als nichtlineares Modell bezeichnet.

### 8.2.1. Nichtlineare Produktionsfunktion

Das grundlegende Modell bleibt beim nichtlinearen Modell erhalten (Gleichung (7.5)). In die Produktionsfunktionen bei zwei Zustandsvariablen ( $y_k$  und  $z_k$ ) werden Sättigungsfunktionen

$$H_{xy}(t) = \frac{(1 + K_{xy})t^h}{t^h + K_{xy}}. \quad (8.7)$$

eingefügt mit der Sättigungskonstante  $K_{xy}$  und dem Hill-Koeffizienten  $h$  mit  $h \geq 1 \in \mathbb{N}$  (siehe Kapitel 2.7).

Die Funktion  $H_{xy}(t)$  erfüllt die für die Normierung der Produktionsfunktionen nötigen Bedingungen  $H_{xy}(0) = 0$ ,  $H_{xy}(1) = 1$  und  $\lim_{x \rightarrow \infty} H_{xy}(x) = 1 + K_{xy}$ . Dies ist notwendig, um die vorher festgelegten Bedingungen für die Produktionsfunktionen zu erfüllen (siehe Kapitel 7.2, Definition 5). Die Halbsättigungskonstante ist bei dieser Funktion  $\bar{x} = \sqrt[h]{K_{xy}}$ , wobei  $K_{xy}$  hier als Sättigungskonstante bezeichnet wird. Mittels der Indizes werden verschiedene Sättigungs-Funktionen definiert. Bei  $H_{xy}(t)$  beeinflusst  $x$  die Produktion von  $y_k$ . Für alle Funktionen wird ein verallgemeinerter Hill-Koeffizient  $h$  verwendet.

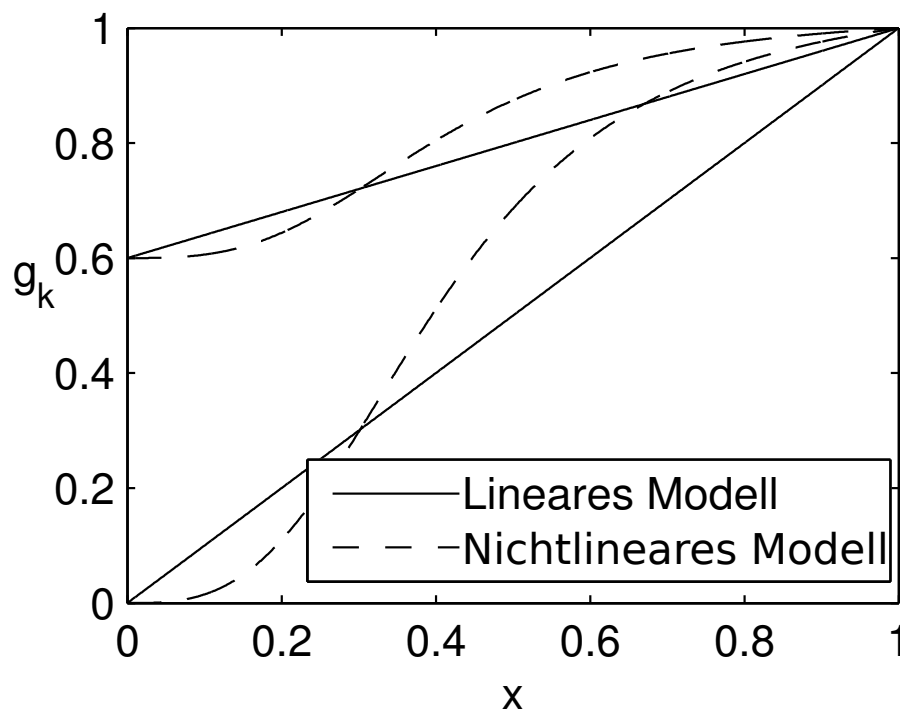
Die Produktionsfunktionen müssen auch mit eingeführten Sättigungsfunktionen die Bedingungen aus Definition 5 erfüllen. Die nichtlinearen Produktionsfunktionen

$$g_k(x) = k + (1 - k)H_{xy}(x), \quad (8.8)$$

$$m(x, y_k) = H_{xz}(x)H_{yz}(y_k) \quad (8.9)$$

beinhalten die Sättigungsfunktionen  $H$ , welche monoton steigend in ihrer Variable sind. Es gilt, dass  $H_{xy}(1) = 1$  ist, somit ist  $H_{xy} \leq 1$ . Um ein geeignetes Sättigungsmanko zu erreichen, sollte die Sättigungskonstante  $K_{xy}$  bei  $0 < K_{xy} \ll 1$  liegen.





**Abbildung 8.3.: Produktionsfunktion des zweiten Regulators ( $g_k(x)$ )**

Die durchgezogene Linien zeigen die lineare Produktionsfunktion mit  $g_k(x) = k + (1 - k)x$  für  $k = 0$  und  $k = 0.6$ . Die gestrichelte Linie stellt die nichtlineare Produktionsfunktion  $g_k(x) = k + (1 - k)H_{xy}(x)$  mit  $H_{xy}(x) = \frac{(1+K_{xy})x^h}{x^h + K_{xy}}$ ,  $K_{xy} = 0.07$ ,  $h = 3$  und  $k = 0$  bzw.  $k = 0.6$  dar.

### 8.2.2. Maximum der basalen Expression

Es existieren Parameter, bei denen die Nutzenbilanz-Funktion  $C(k)$  maximal ist und es gilt, dass die basale Expression in  $k \in ]0, 1[$  liegt, also eine streng maximale basale Expression existiert.

**Satz 8.** *Es existieren nichtlineare Produktionsfunktionen (Gleichung (8.8) und (8.9)), so dass die Nutzenbilanz-Funktion  $C(k)$  (Gleichung (7.8)) des FFL-Modells ein strenges Optimum in  $k \in ]0, 1[$  besitzt.*

**Beweis.** Die Ableitung der Nutzenbilanz-Funktion  $C(k)$  nach  $k$  ist:

$$\frac{dC(k)}{dk} = \dot{C}(k) = \frac{1}{T} \int_0^T s_x \frac{\partial z_k}{\partial k} - \eta_y \frac{\partial g_k(x)}{\partial k} - \eta_z \frac{\partial m(x, y_k)}{\partial k} dt. \quad (8.10)$$

Zur Vereinfachung wird festgelegt, dass

$$\gamma(x) = \frac{\partial g_k(x)}{\partial k} \geq 0, \quad (8.11)$$

$$m_y(x, y_k) = \frac{\partial m(x, y_k)}{\partial y_k}, \quad (8.12)$$

ist.

Mit dem Operator  $L$  kann

$$\frac{\partial m(x, y_k)}{\partial k} = m_y(x, y_k) L[g_k(x)] \gamma, \quad (8.13)$$

$$\frac{\partial y_k}{\partial k} = L[g_k(x)] \gamma, \quad (8.14)$$

$$\frac{\partial z_k}{\partial k} = L[m(x, y_k)] m_y(x, y_k) L[g_k(x)] \gamma. \quad (8.15)$$

berechnet werden.

Die Gleichungen (8.11) bis (8.15) können in die Ableitung der Nutzenbilanz-

Funktion  $C'(k)$  (Gleichung (8.10)) eingesetzt werden:

$$\begin{aligned}\frac{dC(k)}{dk} &= \frac{1}{T} \int_0^T s_x L[m(x, y_k)] m_y(x, y_k) L[g_k(x)] \gamma - \dots \\ &\quad - \eta_y \gamma(x) - \eta_z m_y(x, y_k) L[g_k(x)] \gamma dt, \\ \frac{dC(k)}{dk} &= \frac{1}{T} \int_0^T \gamma \left( (s_x L[m(x, y_k)] - \eta_z) m_y(x, y_k) L[g_k(x)] - \eta_y \right) dt, \\ \frac{dC(k)}{dk} &= \frac{1}{T} \int_0^T s_x z y_k m_y(x, y_k) \gamma - \eta_z y_k m_y(x, y_k) \gamma - \eta_y \gamma dt.\end{aligned}$$

oder vereinfacht

$$\frac{dC(k)}{dk} = \frac{1}{T} \int_0^T \gamma \left( (s_x z - \eta_z) y_k m_y(x, y_k) - \eta_y \right) dt. \quad (8.16)$$

Notwendige Bedingung für ein Maximum ist, dass die Ableitung gleich Null ist. Steigt die Funktion zunächst (Ableitung größer Null) und sinkt dann (Ableitung kleiner Null) und ist  $C(k)$  eine stetige Funktion, muss ein Maximum zwischen dem Wachstum und dem Absinken vorhanden sein. Da  $k \in [0, 1]$  ist, werden

$$\frac{dC}{dk}(0) \quad \text{und} \quad \frac{dC}{dk}(1)$$

betrachtet. Gesucht werden also derartige Parameter, dass  $C'(0) > 0$  und  $C'(1) < 0$  ist.

Die Parameter  $m_y(x, y_k)$ ,  $\eta_y$  und  $\eta_z$  haben Einfluss auf die Steigung der Nutzenbilanz. Zunächst wird die Produktionsfunktion  $m_y(x, y_k)$  betrachtet. Zu zeigen ist, dass für  $k = 1$  die Produktionsfunktion sehr klein ist und für  $k = 0$  diese Produktionsfunktion sehr groß ist.

Die Produktionsfunktion der Zielvariable  $m(x, y_k)$  soll zunächst für  $k = 1$  derart gewählt werden, dass  $m_y(x, 1) = m_y(x, y_1) > 0$  sehr klein ist. Die Produktionsfunktion

$$m(x, y_k) = H_{xz}(x)H_{yz}(y_k) = \frac{(1 + K_{xz})x^h}{x^h + K_{xz}} \frac{(1 + K_{yz})y_k^h}{y_k^h + K_{yz}}$$

kann nach  $y_k$  abgeleitet werden.  $H_{xz}(x)$  ist unabhängig von  $y_k$ , somit ist

$$m_y(x, y_k) = H_{xz}(x) \frac{\partial H_{yz}(y_k)}{\partial y_k} = H_{xz}(x) \left( \frac{(1 + K_{yz})y_k^h h}{y_k(y_k^h + K_{yz})} - \frac{(1 + K_{yz})(y_k^h)^2 h}{(y_k^h + K_{yz})^2 y_k} \right)$$

Ist  $k = 1$  gilt auch  $y = 1$ , somit gilt

$$m_y(x, 1) = H_{xz}(x) \left( \frac{(1 + K_{yz})1^h h}{1(1^h + K_{yz})} - \frac{(1 + K_{yz})(1^h)^2 h}{(1^h + K_{yz})^2 1} \right) = H_{xz}(x) \left( h - \frac{h}{1 + K_{yz}} \right)$$

Für genügend kleine  $K_{yz}$  ist  $h - (h/(1 + K_{yz}))$  sehr klein und somit kann  $m_y(x, 1)$  sehr klein sein.

Die Produktionsfunktion  $m_y(x, y_k)$  muss, wenn die basale Expression gleich Null  $k = 0$ , genügend groß sein. Hier gilt, dass

$$m_y(x, y_0) = H_{xz}(x) \left( \frac{(1 + K_{yz})y_0^h h}{y_0(y_0^h + K_{yz})} - \frac{(1 + K_{yz})(y_0^h)^2 h}{(y_0^h + K_{yz})^2 y_0} \right)$$

ist mit  $y_0 \in [0, 1]$ . Somit muss der Hill-Koeffizient  $h$  und  $H_{xz}(x)$  groß genug sein, damit  $m_y(x, y_0)$  groß genug wird mit genügend kleinen  $K_{yz}$ .

Der Term  $(s_{xz} - \eta_z)y_k m_y(x, y_k)$  ist sehr groß für  $k = 0$  und für  $k = 1$  sehr klein, wenn gilt, dass  $m_y(x, y_0) > 0$  sehr groß ist für alle Werte von  $y_0$ ,  $m_y(x, 1) = m_y(x, y_1) > 0$  sehr klein ist und  $\eta_z$  klein genug und somit  $s_{xz} - \eta_z$  positiv für alle  $k$  ist. Negativen Einfluss auf die Nutzenbilanz-Funktion hat somit nur der Teil  $\eta_y$ . Der Kostenfaktor  $\eta_y$  kann derart gewählt werden,

dass

$$C'(0) = \frac{1}{T} \int_0^T \gamma((s_x z - \eta_z) y_0 m_y(x, y_0) - \eta_y) dt > 0$$

$$C'(1) = \frac{1}{T} \int_0^T \gamma((s_x z - \eta_z) y_1 m_y(x, y_1) - \eta_y) dt < 0$$

ist. Somit kommt es zu einem strengen Maximum der Nutzenbilanz-Funktion  $C(k)$  in  $k \in ]0, 1[$ , für entsprechend gewählte  $\eta_y$ ,  $\eta_z$  und  $m_y(x, y_0)$ .  $\square$

### 8.2.3. Grenzbereiche

Auch im Grenzbereich bei sehr schnell wechselnden Signalen lässt sich ebenfalls ein Optimum der Nutzenbilanz-Funktion berechnen. Gegeben sind schnell alternierende Signale  $\bar{s} = s_x(\bar{T})$  mit  $\bar{T} = \bar{T}_{i+1} - \bar{T}_i \rightarrow 0, \forall i$ . Der Auf- und Abbau der Variablen bei schnell alternierende Signale wird reduziert, daher tendieren die Lösungen der Differentialgleichungen (Gleichung (7.5)) gegen konstante Funktionen, nämlich zu den Gleichgewichtsfunktionen der Differentialgleichungen.

$$x \rightarrow \bar{s},$$

$$y_k \rightarrow g_k(\bar{s}),$$

$$z_k \rightarrow m(\bar{s}, g_k(\bar{s})).$$

Die Nutzenbilanz-Funktion tendiert dann zu

$$\begin{aligned} \lim_{\bar{T}_{i+1}-\bar{T}_i \rightarrow 0} C(k) &= \lim_{\bar{T}_{i+1}-\bar{T}_i \rightarrow 0} \frac{1}{\bar{T}_{i+1}-\bar{T}_i} \int_{\bar{T}_i}^{\bar{T}_{i+1}} s_x z_k - \eta_y g_k(x) - \eta_z m(x, y_k) dt, \\ &= \bar{s} m(\bar{s}, g_k(\bar{s})) - \eta_y g_k(\bar{s}) - \eta_z m(\bar{s}, g_k(\bar{s})) \end{aligned}$$

und die Ableitung der Nutzenbilanz-Funktion zu

$$\begin{aligned} \lim_{\bar{T}_{i+1}-\bar{T}_i \rightarrow 0} \frac{dC(k)}{dk} &= \lim_{\bar{T}_{i+1}-\bar{T}_i \rightarrow 0} \frac{1}{\bar{T}_{i+1}-\bar{T}_i} \int_{\bar{T}_i}^{\bar{T}_{i+1}} \gamma((s_x z - \eta_z) y_k m_y(x, y_k) - \eta_y) dt, \\ &= \gamma(\bar{s}) ([\bar{s} \cdot m(\bar{s}, g_k(\bar{s})) - \eta_z] g_k(\bar{s}) m_y(\bar{s}, g_k(\bar{s})) - \eta_y). \end{aligned}$$

Wenn gilt, dass  $\eta_z < \bar{s} m(\bar{s}, g_0(\bar{s}))$  ist, also der Term  $p_1 = \bar{s} \cdot m(\bar{s}, g_k(\bar{s})) - \eta_z > 0$  ist, ist somit

$$\lim_{\bar{T}_{i+1}-\bar{T}_i \rightarrow 0} \frac{dC(k)}{dk} = \gamma(\bar{s}) (p_1 g_k(\bar{s}) m_y(\bar{s}, g_k(\bar{s})) - \eta_y).$$

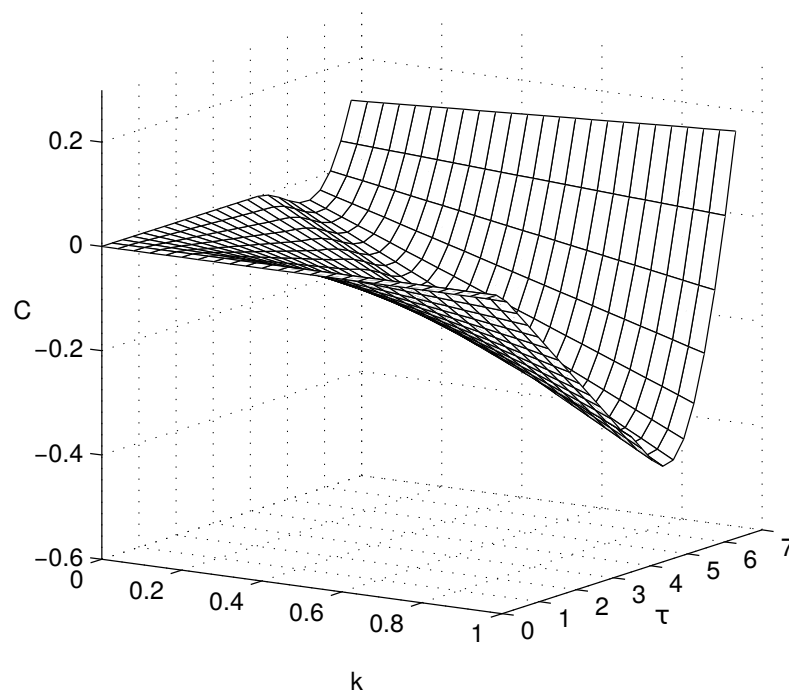
Wenn  $m_y$  derart für  $k = 0$  und  $k = 1$  gewählt wird, dass  $m_y(\bar{s}, 1) \ll m_y(\bar{s}, g_0(\bar{s}))$  gilt, kann  $\eta_y$  so gewählt werden, dass

$$[\bar{s} m(\bar{s}, 1) - \eta_z] m_y(\bar{s}, 1) < [\bar{s} m(\bar{s}, g_0(\bar{s})) - \eta_z] g_0(\bar{s}) m_y(\bar{s}, g_0(\bar{s})).$$

gilt. Dann ist die Ableitung der Nutzenbilanz-Funktion  $C(k)$  auch für schnell alternierende Signale für  $k = 0$  positiv  $C'(0) > 0$  und für  $k = 1$  negativ  $C'(1) < 0$ . Ein strenges Maximum der Nutzenbilanz-Funktion  $C(k)$  für schnell alternierende Signale existiert, da  $C(k)$  auch bei schnellen Wechsel des Signals stetig definiert ist.

Zwar existiert auch ein strenges Maximum der basalen Expression, auch für

schnell alternierende Signale, doch hat genau wie im linearen FFL-Modell auch das nichtlineare Modell eine negativen momentanen Nutzenbilanz  $c_k$  direkt nach dem ersten Wechsel des Signals  $T_1$ . Daher konnte dort zwar bewiesen werden, dass die Nutzenbilanz maximal  $\max(C(k))$  ist, aber es konnte nicht bewiesen werden, dass die Nutzenbilanz abhängig von der Zeit wächst. Daher ist nach Definition 9 das Modell nicht evolutionär vorteilhaft.



**Abbildung 8.4.: Die Nutzenbilanz direkt nach dem ersten Wechsel des Signals innerhalb des nichtlinearen Modells**

Die Nutzenbilanz des linearen Modells  $C(k)$  wird direkt nach dem Wechsel des Signals bei  $\tau = t - T_1$ . Die Kosten-Faktoren wurden mit  $\eta_y = \eta_z = 0.1$  festgelegt. Der Hill-Koeffizient wurde hier mit  $h = 3$  und die Sättigungskonstanten mit  $K_{xy} = K_{xz} = K_{yz} = 0.1$  festgelegt. Die Nutzenbilanz-Funktion zeigt Negativität für kleine  $\tau$ . Bei größeren  $\tau$  wächst der momentane Nutzen mehr als die momentanen Kosten. Somit wächst auch  $C(k)$ .

In Abbildung 8.4 ist die Nutzenbilanz  $C(k)$  direkt nach dem ersten Wechsel des Signals dargestellt. Dies ist für das lineare Modell bereits in Abbildung 8.1 dargestellt. Die Nutzenbilanz ist ab dem ersten Wechsel mit  $\tau = t - T_1$

und dem Startwert  $C(k)_{t=T_1} = 0$  dargestellt. Hierbei ist  $\tau$  die Zeit nach der ersten Schaltung. Direkt nach dem Signalwechsel sinkt die Nutzenbilanz. Das Modell ist trotz des optimalen  $k$  (basale Expression) nicht evolutionär vorteilhaft (siehe Definition 9). Eine evolutionär vorteilhafte Erklärung der FFL mit basaler Expression im zweiten Regulator wird in dem nächsten Abschnitt durch Erweiterung des Modells gesucht.

#### 8.2.4. Simulation mit dem nichtlinearen Modells

Die Simulationen mit dem nichtlinearen Modell in Abbildung 8.5 ist mit den gleichen Parametern und dem gleichen Signal wie die Simulation des linearen Modells aus Abbildung 8.2 und der Sättigungskonstanten der nichtlinearen Produktionsfunktionen und der Hill-Koeffizient

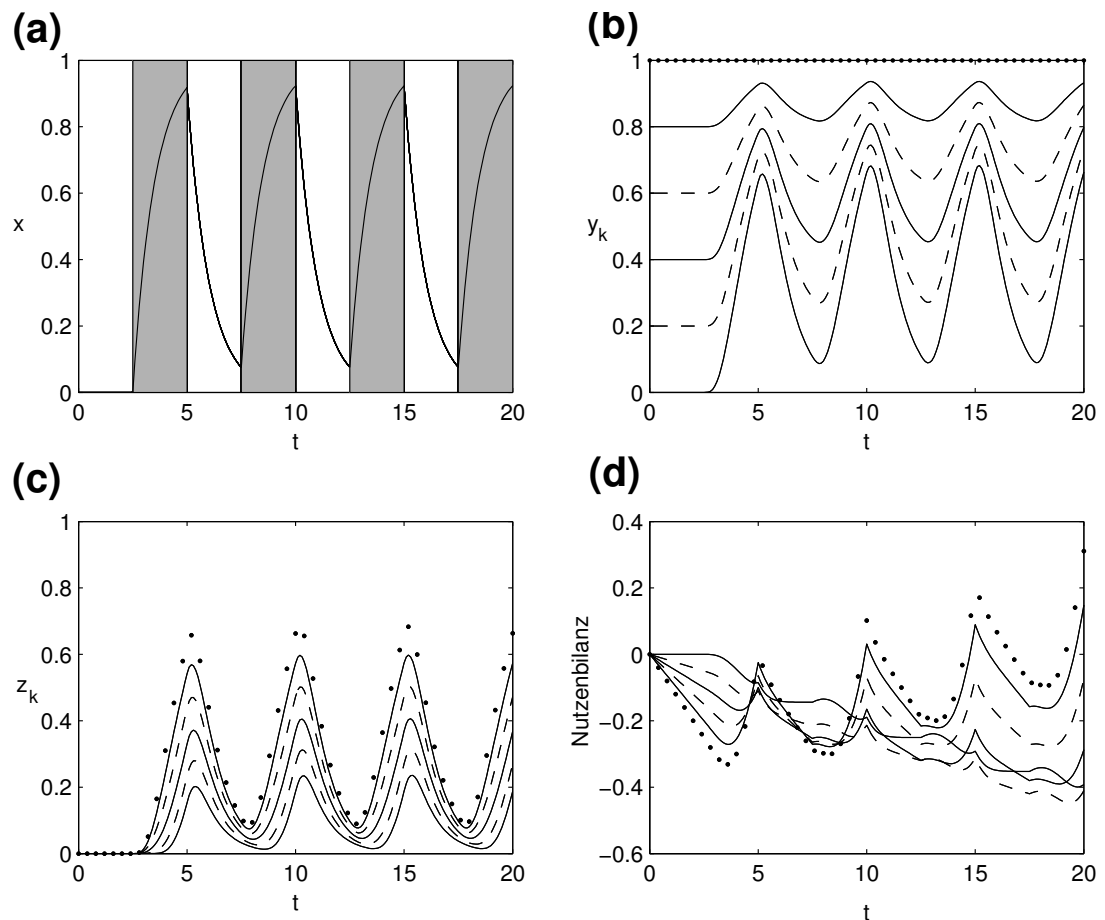
$$K_{xy} = K_{xz} = K_{yz} = 0.07,$$

$$h = 3.$$

durchgeführt.

Die Simulation mit dem nichtlinearen Modell (siehe Abbildung 8.5) sieht auf den ersten Blick zunächst der Simulation des linearen Modells (siehe Abbildung 8.2) sehr ähnlich. Der erste Regulator wird genau wie im linearen Modell exprimiert. Die Zustandsvariable  $x$  unterscheidet sich in den Modellen nicht. Dahingegen unterscheidet sich der zweiten Regulator  $y_k$  in den Modellen. So ist der Unterschied des Auf- und Abbaus von  $y_k$  je nach der basalen Expression  $k$  in dem nichtlinearen Modell größer als in dem linearen Modell. Dies hängt von den eingegebenen Hill-Koeffizienten  $h$  und der Sättigungskonstante ab.





**Abbildung 8.5.: Simulationsergebnis des nichtlinearen Modells.**

Die Phasen mit Signal werden durch einen grauen Hintergrund bei der Zustandsvariable  $x$  dargestellt (a). In (b) wird der zweite Regulator  $y_k$  für die basalen Expression  $k = 0, 0.2, 0.4, 0.6, 0.8$  (durchgängige, gestrichelte, durchgängige, gestrichelte und durchgängige Linien) und  $k = 1$  (gepunktete Linie) gezeigt. Die Darstellung der Zielvariable  $z_k$  befinden sich in der Abbildung (c). Kosten-Faktoren der Nutzenbilanz  $C(k)$  wurden mit  $\eta_y = \eta_z = 0.1$  gesetzt (d). Die Sättigungskonstanten sind alle als  $K_{xy} = K_{xz} = K_{yz} = 0.07$  festgelegt und alle Hill-Koeffizienten mit  $h = 3$  gewählt.

Die veränderten Produktionsfunktionen von  $y_k$  und  $z_k$  wirken sich auf die Nutzenbilanz-Funktion  $C(k)$  aus. Bei dem linearen Modell erkennt man, dass entweder das Modell mit fehlender basaler Expression ( $k = 0$ ) oder mit vollständiger basaler Expression ( $k = 1$ ) die maximale Nutzenbilanz-Funktion besitzt (siehe Kapitel 8.1.3). Bei dem nichtlinearen Modell kann

auch ein Modell mit mittlerer basaler Expression eine maximale Nutzenbilanz haben.

Innerhalb des nichtlinearen Modell gibt es bedingt eine streng maximale basale Expression  $k$ . Diese Bedingungen sind, dass nur Parametern  $\eta_y$ ,  $\eta_z$  und  $m_y(x, y_0)$  in einem geeigneten Wertebereich zu einer streng optimale basalen Expression führen. Jedoch sinkt auch hier die Nutzenbilanz bei Signalwechsel. Das Modell ist nicht evolutionär vorteilhaft.

## 8.3. Modell mit Speicher

In den Kapiteln 8.1 und 8.2 wurden zwei Modelle für die FFL mit basaler Expression vorgestellt. In Kapitel 8.2 wurde bereits gezeigt, dass ein nichtlineares Modell die basale Expression erklären kann. Es gibt folglich Parameter, bei denen es sinnvoll für ein Bakterium ist, den zweiten Regulator  $y_k$  basal zu exprimieren. Es wird ein Modell gesucht, das genauso den Nutzen der basale Expression erläutern, aber auch evolutionäre vorteilhaft ist, daher auch die momentane Nutzenbilanz positiv ist.

### 8.3.1. Notwendigkeit der Erweiterung

Innerhalb des linearen und des nichtlinearen Modells werden direkt nach einem Wechsel mehr Kosten als Nutzen produziert, wie in Abbildung 8.1 und 8.4 dargestellt. Die Nutzenbilanz-Funktion  $C(k)$  sinkt und somit ist die momentane Nutzenbilanz negativ. Zu dem Zeitpunkt des Signalwechsels zeigen die beiden bisherigen Modelle keinen evolutionären Vorteil (siehe Definition 9). Gesucht wird ein Modell, welches zu der Modellfamilie mit den in Kapitel 7 aufgestellten Bedingungen gehört und gleichzeitig alle Bedingungen

des evolutionären Vorteil erfüllt. Um diese Bedingungen zu erfüllen, wird ein Energiespeicher eingeführt.

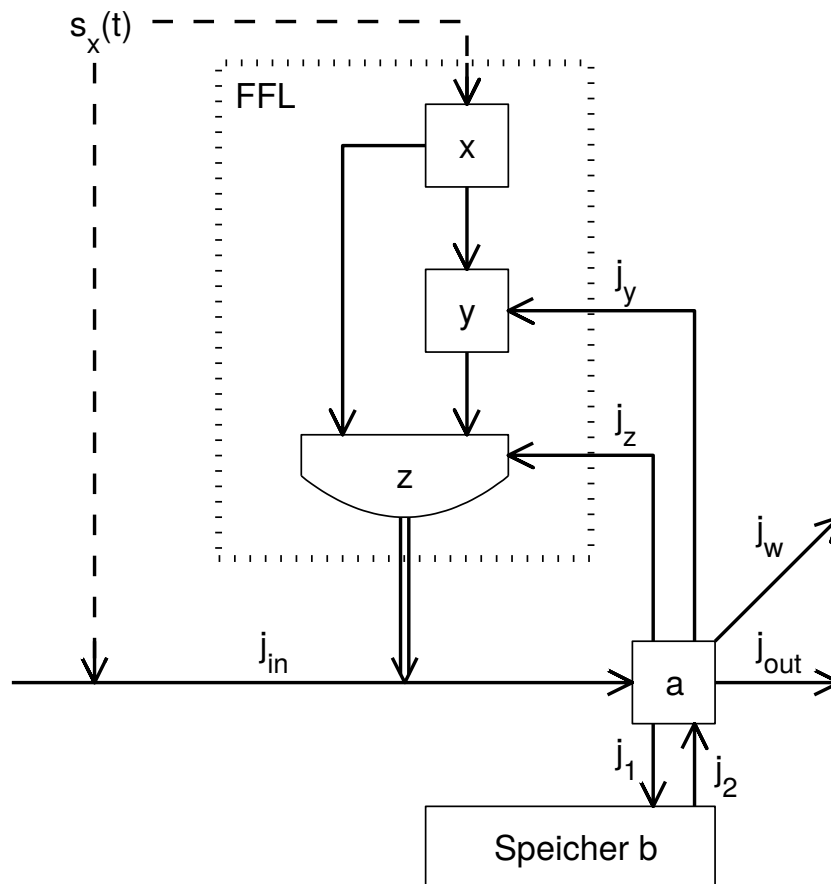
### 8.3.2. Energiespeicher

Jedes Bakterium hat bei kurzfristigem Energiebedarf Energiespeicher zur Verfügung, kann also durch gespeichert Energie diese negative Bilanz ausgleichen. Vielen Bakterien, einschließlich *P. aeruginosa*, können Energie zum Beispiel als Polyhydroxyalkanoate (PHA) und Polyphosphosphate speichern, um schwankenden Umweltbedingungen zu kompensieren (Chan et al., 2006, Achbergerová and Nahálka, 2011). Im Gegensatz dazu können ATP, aber auch andere Moleküle wie Acetyl-CoA und NADH Energie für den unmittelbaren Gebrauch liefern.

Ein Energiespeicher kann benötigte Energie liefern und somit energetische Kosten abdecken. Eine negative Nutzenbilanz kann so von gespeicherter Energie abgefangen werden. Der Energiespeicher wird mit einem Energieverteiler eingeführt. Die Nutzenbilanz muss neu definiert werden. Hier werden Nutzen und Kosten von dem Energiespeicher zusammengefasst. Das Bakterium kann aus einem Energieüberschuss den Nutzen des Wachstums ziehen. Anstelle der Nutzenbilanz soll Wachstum optimiert werden. In Abbildung 8.6 wird das erweiterte System bildlich dargestellt.

Das Grundmodell liefern die Differentialgleichungen zu den Zustandsvariablen  $x$ ,  $y_k$  und  $z_k$  (siehe Gleichungen (7.5)). Neu definiert werden die Zustandsvariablen für den Energieverteiler ( $a$ ), den Energiespeicher ( $b$ ) und die Fitness ( $w$ ). Die Ableitung der Fitness nach der Zeit wird als Wachstum ( $\dot{w}$ ) bezeichnet. Der Energiespeicher  $b$  kann Energie für die Produktion der Zustandsvariablen  $y_k$  und  $z_k$  speichern. Der Energieverteiler  $a$  kann die benötigten Energien zuordnen. Aus dem Energiespeicher fließt die Energie mit der Rate  $j_2$  in den Verteiler, wenn dieser Energie benötigt. Mit  $j_1$

fließt überschüssige Energie aus dem Verteiler in den Speicher zurück. Immer wenn der Energieverteiler Energie benötigt, bezieht er diese aus dem Energiespeicher. Wird genügend Energie von dem Modell produziert, wird Energie gespeichert.



**Abbildung 8.6.: Das Modell der FFL mit Speicher**

beinhaltet das einfach FFL-Modell und einen Energie-Speicher  $b_k$  und einen Energie-Verteiler  $a_k$ . Das  $(x, y_k, z_k)$ -System ist als Sub-Modell (durch die gestrichelte Linie markiert) eingebettet. Durchgängige einzelne Linien stellen den Energiefluss dar, Signale sind gestrichelte Linien, die doppelte Linien zeigt den Einfluss von  $z_k$  auf den Energie-Verteiler  $a_k$ . Die Energie fließt in den Speicher  $b_k$  mit  $j_1$  und aus dem Speicher mit der konstanten Rate  $j_2$ . Für die Produktion von  $y_k$  und  $z_k$  fließt Energie mit den konstanten Raten  $j_y$  und  $j_z$ . Der Grundbedarf an Energie fließt mit  $j_{out}$  aus dem System und das Wachstum wird mit  $j_w$  definiert.

Es kostet Energie, den zweiten Regulator  $y_k$  und die Zielvariable  $z_k$  zu pro-

duzieren. In dem Modell ohne Speicher wurden diese Kosten von den Nutzen abgezogen und man erhielt die Nutzenbilanz. In dem neu aufgestellten Modell mit Speicher können diese Kosten mit den Raten  $j_y$  und  $j_z$  von dem Energieverteiler abgedeckt werden.

Der Nutzen aus  $z_k$  existiert bei Anwesenheit des Signals  $s_x$ , genau so wie in dem Modell ohne Speicher. Der energetische Nutzen aus  $z_k$  geht hier mit dem Parameter  $j_{in}$  in den Verteiler und Speicher ein. Dieser Nutzen kann somit den Energiespeicher wieder auffüllen, auf dem bei Produktion von  $y_k$  und  $z_k$  zurückgegriffen werden kann.

Bei Abwesenheit des Signals wird Energie durch andere Wege produziert, im Falle der Atmungsregulation von *P. aeruginosa* z. B. durch Oxidation. Ist also das Signal  $s_x = 0$ , so wird ebenfalls Energie produziert, welche auch in den Verteiler fließt.

Das Bakterium benötigt durchgängig Energie für seinen Stoffwechsel. Diese Energie geht mit  $j_{out}$  aus dem Verteiler in den Grundmetabolismus. Diese Energie für den Grundmetabolismus wird dauerhaft benötigt und fließt somit konstant aus dem Verteiler.

Alle Werte aus der Nutzenbilanz des Modells ohne Speicher fließen in diesem Modell mit Speicher in den Verteiler. Hat der Verteiler Energie, kann diese in das Wachstum  $\dot{w}$  der Bakterien fließen. Ein Wachstum existiert somit nur bei Energieüberschuss. In diesem Fall wird mit einer Rate von  $j_w$  Energie in das Wachstum gesteckt. Das integrierte Wachstum wird als Fitness definiert. Das Modell wird, wie in Definition 9, als evolutionär vorteilhaft definiert, wenn sowohl das Wachstum  $\dot{w}$  als auch die Fitness  $w$  positiv sind als auch die Fitness in Abhängigkeit von der basalen Expression streng maximal ist.

Sowohl die Konzentration an Energie in dem Energieverteiler, die in dem Energiespeicher als auch die Fitness hängen von der basalen Expression  $k$

ab. Daher werden sie mit  $a_k$ ,  $b_k$  und  $w_k$  betitelt.

Insgesamt kommt man somit zu einem erweitertem System

$$\begin{aligned}
 \dot{x} &= s_x - x, \\
 \dot{y}_k &= g_k(x) - y_k, \\
 \dot{z}_k &= m(x, y_k) - z_k, \\
 \dot{a}_k &= j_{\text{in}} - j_1 + j_2 - j_y - j_z - j_{\text{out}} - j_w, \\
 \dot{b}_k &= j_1 - j_2, \\
 \dot{w}_k &= j_w,
 \end{aligned} \tag{8.17}$$

welches die für den Energiefluss zuständigen Parameter  $j_{\text{in}}$ ,  $j_{\text{out}}$ ,  $j_y$ ,  $j_z$ ,  $j_1$ ,  $j_2$  und  $j_w$  besitzt.

Solange der Energiespeicher  $b_k$  nicht voll genug ist, also weniger beinhaltet als eine gesetzte Größe von  $b_{\text{set}}$ , wird dieser mit Energie aus dem Verteiler aufgefüllt  $b_k < b_{\text{set}}$  (Langemann and Rehberg, 2010). Es wird der Einfachheit halber angenommen, dass der Energiefluss von dem Speicher abhängig von dem Unterschied zwischen der Speicherfülle  $b_k$  und der gesetzten Größe  $b_{\text{set}}$  ist,  $j_1 \sim b_{\text{set}} - b_k$ . Durch  $j_1$  ist die Rate des Energieflusses definiert. Durch Normierung wird  $b_{\text{set}}$  auf eins gesetzt ( $b_{\text{set}} = 1$ ). Der Energiespeicher gibt Energie an den Verteiler ab, wenn der Verteiler negativ ist, also wenn er Energie braucht. Es kann nur Energie vom Speicher abgegeben werden, wenn dieser genügend Energie beinhaltet. Der Energiefluss besteht bei  $b_k > 0$ , also wenn der Speicher Energie hat. Der positive Anteil der Energie im Speicher wird definiert durch  $b_k^+ = (b_k + |b_k|)/2 = \max(0, b_k)$ . Es wird nur dann Energie benötigt, wenn der Verteiler  $a_k$  eine negative Energiebilanz besitzt, also wird nur Energie aus dem Speicher gezogen, wenn  $a_k < 0$  ist. Der negative Anteil des Verteilers ist gegeben durch  $a_k^- = (|a_k| - a_k)/2 = |\min(a_k, 0)|$ . Der Energiefluss aus dem Speicher zu dem Verteiler wird al-

so proportional zu diesen beiden Größen sein,  $j_2 \sim a_k^- b_k^+$ . Die Geschwindigkeit des Energieflusses wird durch den Proportionalitätsfaktor  $\beta$  dargestellt.

Das Bakterium wächst nur dann mit einer Rate  $\alpha$ , wenn die Energiebilanz des Verteilers positiv, also wenn  $a_k > 0$  ist ( $a_k^+ = (a_k + |a_k|)/2 = \max(a_k, 0)$ ). Die Parameter

$$\begin{aligned}
 j_{\text{in}} &= 1 - s_x + s_x z_k, \\
 j_{\text{out}} &= \text{const}, \\
 j_y &= \eta_y g_k(x), \\
 j_z &= \eta_z m(x, y_k), \\
 j_1 &= 1 - b_k, \\
 j_2 &= \beta a_k^- b_k^+, \\
 j_w &= \alpha a_k^+,
 \end{aligned} \tag{8.18}$$

zeigen die Energieflüsse des Systems aus Gleichung (8.17).

Das extreme Szenario, dass die Signale sehr kurz hintereinander wechseln, wurde für das Modell ohne Speicher bereits erläutert (siehe Kapitel 8.2.3). Dabei wurde berechnet, dass bei extrem schneller Schaltung die Zustandsvariablen zu Konstanten  $x \rightarrow \bar{s}$ ,  $y_k \rightarrow g_k(\bar{s})$  und  $z_k \rightarrow m(\bar{s}, g_k(\bar{s}))$  streben. Auch die Zustandsvariablen  $a_k$  und  $b_k$  werden zu Konstanten. Es wäre dann  $j_1 = j_2$  und  $j_w = j_{\text{in}} - j_y - j_z - j_{\text{out}}$ . Das Wachstum  $\dot{w}_k$  kann für  $\bar{s}$  als konstant beschrieben werden,

$$\dot{w}_k = 1 - \bar{s} + \bar{s} z_k - \eta_y g_k(\bar{s}) - \eta_z m(\bar{s}, g_k(\bar{s})) = \text{const}.$$

Selbst für extreme Bedingungen ist das Wachstum  $\dot{w}_k$ , bzw die Fitness  $w_k$  berechenbar.

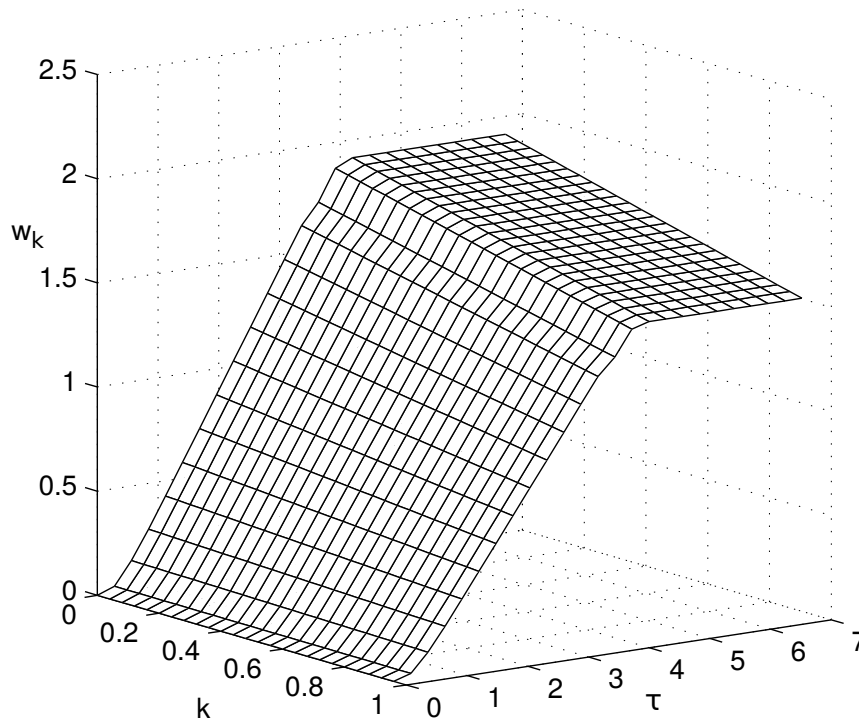
### 8.3.3. Lineares Modell mit Speicher

In dem Modell mit Speicher können ebenfalls unterschiedliche Produktionsfunktionen für den zweiten Regulator  $y_k$  und der Zielvariablen  $z_k$  eingefügt werden. Zunächst werden die linearen Produktionsfunktionen

$$g_k(x) = k + (1 - k)x,$$

$$m(x, y_k) = x \cdot y_k,$$

bekannt aus Kapitel 8.1, genutzt.



**Abbildung 8.7.:** Die Fitness  $w_k$  des linearen Modells mit Energiespeicher für variierende basale Expressionen direkt nach einem Signalwechsel  $T_1$

Hier wird  $\tau = t - T_1$  genutzt mit  $T_1$  als Zeitpunkt des Wechsels und  $\tau$  Zeit nach dem Wechsel. Die Kosten-Faktoren werden festgelegt als  $\eta_y = \eta_z = 0.1$  und die anderen Parameter mit  $\alpha = 4$ ,  $\beta = 10$  und  $j_{\text{out}} = 0.5$ . Die Fitness zeigt über die gesamte Zeit keine Negativität.



Durch das Energiespeichersystem kommt bei einem Signalwechsel zu keiner negativen momentanen Nutzenbilanz. Direkt nachdem dem ersten Wechsel des Signals bei  $t = T_1 + \tau$ , mit  $T_1$  als Zeitpunkt des Wechsels und  $\tau$  als Zeit danach, zeigt die momentane Fitness  $w_k$  keine Negativität und nur Steigung, also das momentane Wachstum  $\dot{w}_k \geq 0$ .

Das Wachstum des linearen FFL-Modell mit einem Energiespeicher ist in Abbildung 8.7 gezeigt, mit Kosten-Faktoren  $\eta_y = \eta_z = 0.1$ ,  $\alpha = 4$ ,  $\beta = 10$  und  $j_{\text{out}} = 0.5$ .

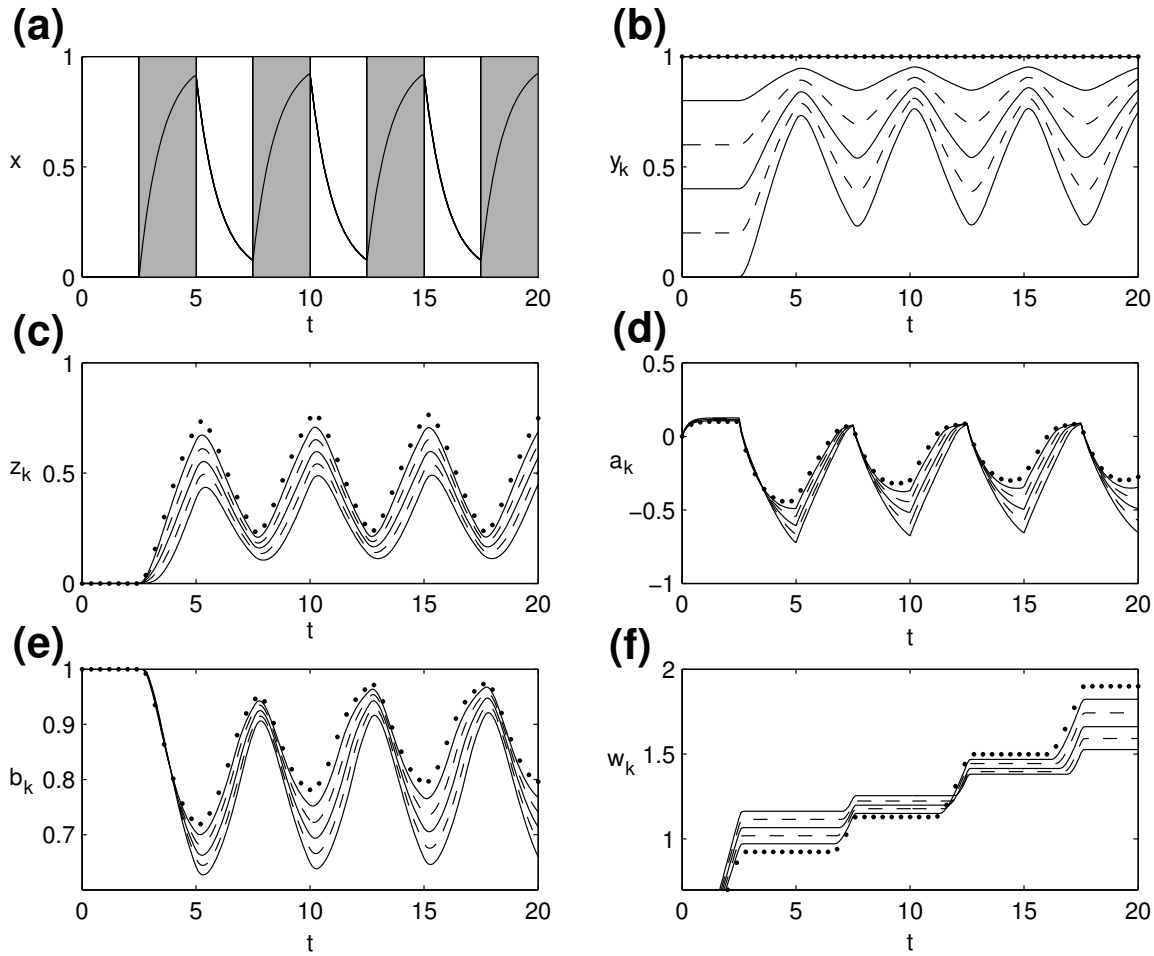
Die Verwendung des FFL-Modells mit Speicher mit linearen Produktionsfunktionen erzeugen ein nur positives Wachstum. Somit ist eine Voraussetzung für evolutionären Vorteil erfüllt. Zu zeigen ist nun, dass das Modell mit Speicher ein streng maximale Fitness abhängig von der basalen Expression  $\max(C(k))$  hat.

### Simulation mit linearem Speichermodell

Innerhalb der Simulation mit dem linearen Modell mit Speicher in Abbildung 8.8 werden die gleichen periodischen Signale angenommen wie in den Modellen ohne Speicher (Gleichung (8.6)).

Die Anfangsbedingungen sind  $x(0) = 0$ ,  $y_k(0) = k$ ,  $z_k(0) = 0$ ,  $a_k(0) = 0$ ,  $b_k(0) = b_{\text{set}} = 1$  und  $w_k(0) = 0$ . Genutzt werden für das Modell aus Gleichung (8.17) mit den dazugehörigen Konstanten aus Gleichung (8.18) die folgenden Parameter:

$$\begin{aligned}\alpha &= 4, \\ j_{\text{out}} &= 0.5, \\ \beta &= 1, \\ \eta_y &= \eta_z = 0.1.\end{aligned}$$

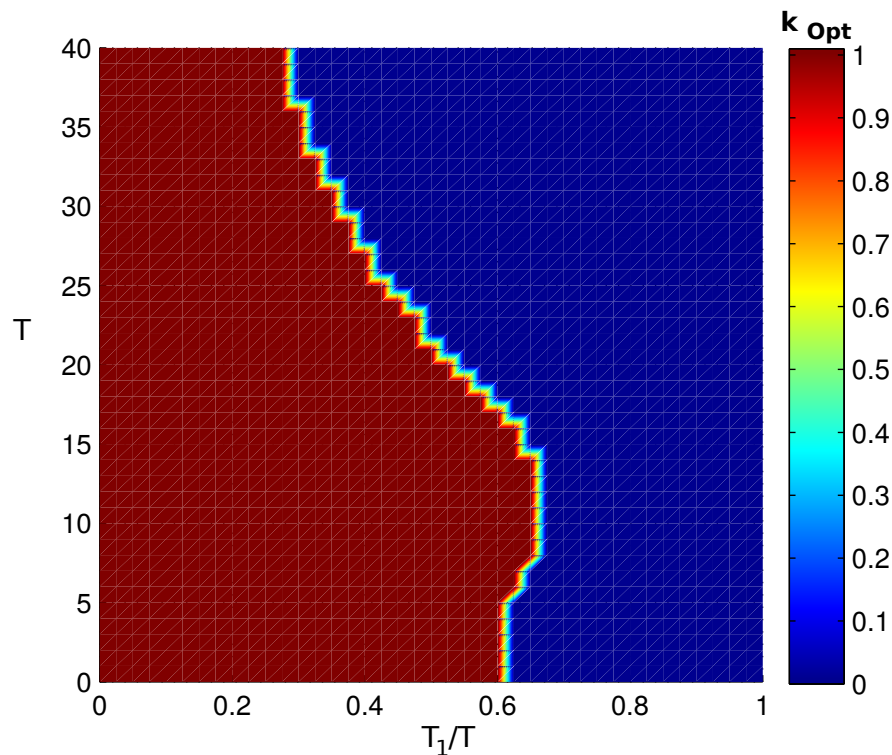


**Abbildung 8.8.: Simulations-Ergebnisse des linearen Modells mit Speicher**

Simulation für  $k = 0, 0.2, 0.4, 0.6, 0.8, 1$  (durchgängige, gestrichelte, durchgängige, gestrichelte, durchgängige, gepunktete Linien). Der erste Regulator  $x$  und die Phasen ohne Signal werden mit grauen Hintergrund in (a) gezeigt. (b) zeigt den Verlauf des zweiten Regulators  $y_k$ . Die Zielvariable  $z_k$  befindet sich in Abbildung (c). (d) und (e) zeigen je den Verteiler- ( $a_k$ ) und den Speicherverlauf ( $b_k$ ), welche abhängig von  $x$ ,  $y_k$  und  $z_k$  steigen und fallen. Die Fitness  $w_k$  (f) steigt monoton. Sobald in dem Verteiler  $a_k$  Energie vorliegt, kommt es zu Wachstumsschüben. Zu allen anderen Zeiten ist die Fitness konstant.

Bei Anwesenheit des Signals wachsen die drei Zustandsvariablen  $x$ ,  $y_k$  und  $z_k$  wie in dem linearen Modell ohne Speicher (siehe Gleichungen (7.5)), dabei hängen die Variablen  $y_k$  und  $z_k$  von der basalen Expression  $k$  des zweiten

Regulators  $y_k$  ab. Je höher die basale Expression, desto höher auch  $y_k$  und  $z_k$ .



**Abbildung 8.9.: Optimale basale Expression des linearen Speichermodells**

Optimale basale Expression ( $k$ ) des linearen Modells mit Speicher, mit  $T$  als Gesamtperiodenlänge und  $T_1/T$  als Anteil ohne Signal. Dunkelblau zeigt ein optimales Wachstum  $w_k$  für keine basale Expression ( $k = 0$ ). Bei dunkelrot liegt ein maximales Wachstum bei voller basaler Expression ( $k = 1$ ). Die Farben von dunkelblau über gelb bis dunkelrot stellen eine basale Expression zwischen null und eins dar ( $k \in ]0, 1[$ ) vor.

Bei Anwesenheit des Signals ( $s_x = 1$ ) wird zunächst viel Energie zur Produktion von  $y_k$  und  $z_k$  benötigt. Die Energiekonzentration im Verteiler  $a_k$  und im Energiespeicher  $b_k$  sinkt. Die Fitness  $w_k$  hingegen stagniert. Ist das Signal nicht mehr vorhanden ( $s_x = 0$ ), wird Energie in dem Verteiler und somit auch in dem Speicher wieder aufgebaut. Verbraucht wird nur noch ein minima-

ler Anteil der Energie für die Produktion der basalen Expression. Erst wenn der Wert des Verteilers  $a_k$  über Null liegt, kommt es auch zu Wachstum, die Fitness wird größer. Um so größer die basale Expression eines Bakterium, um so früher wächst die Fitness nach Wechsel des Signals. Jedoch hat eine kleinere basale Expression auch kleinere Kosten. Schon nach vier Perioden ist das Bakterium mit basaler Expression besser aufgestellt als das ohne. Das Optimum der Fitness befindet sich bei dem linearen Modell mit Speicher in der gezeigten Simulation bei  $k = 1$ .

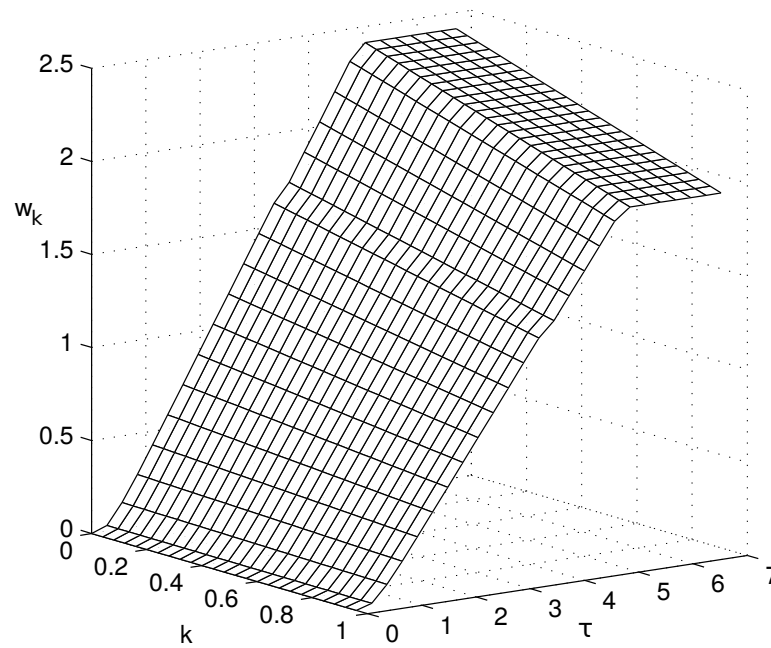
Bei dem linearen Modell mit Speicher lässt sich in der Simulation (Abbildung 8.8) erkennen, dass entweder das Modell mit voller basaler Expression oder das Modell ohne basale Expression die größte Fitness besitzt. Dies ist nicht nur in der gezeigten Simulation der Fall. Für jede Umgebung wird das lineare Modell mit Speicher in Abbildung 8.9 dargestellt. Die optimale basale Expression  $k$  wird zu der maximalen Fitness  $w_k$  berechnet. Es werden periodischen Signal mit der maximalen Periodenlänge von  $T = 40$  genutzt. Zu jeder Periodenlänge von  $T \in ]0, 40]$  und Anteil der Zeit ohne Signal ( $T_1/T \in [0, 1]$ ) wird das optimale  $k$  berechnet. Dunkelblaue Stellen zeigen, dass hier ein maximales Wachstum für fehlende basale Expression vorliegt ( $k = 0$ ). Bei dunkelroter Farbe ist das optimale Wachstum für maximale basale Expression vorhanden ( $k = 1$ ).

Das lineare Modell mit Speicher zeigt keine optimale Fitness für eine basale Expression zwischen null und eins  $k \in ]0, 1[$ , unabhängig von der Umgebung. Je nach Umgebung wird jedoch eine maximale Fitness für keine basale oder volle basale Expression  $k \in \{0, 1\}$  erreicht.

#### 8.3.4. Nichtlineares Modell mit Speicher

Es wurde bereits in Kapitel 8.2.2 gezeigt, dass ein Modell ohne Speicher, aber mit nichtlinearen Produktionsfunktionen eine maximale Nutzenbilanz-

Funktion hat für ein  $k \in ]0, 1[$ . Das Modell mit Speicher (Kapitel 8.3.3) und das Modell mit nichtlinearen Produktionsfunktionen (Kapitel 8.2) werden kombiniert, um ein evolutionär vorteilhaftes Modell zu erreichen.



**Abbildung 8.10.: Die Fitness  $w_k$  direkt nach dem ersten Wechsel des Signals innerhalb des nichtlinearen Modells mit Speicher**

Der Hill-Koeffizient wird hier mit  $h = 3$  und die Sättigungskonstanten mit  $K = 0.1$  festgelegt. Es wird  $\tau = t - T_1$  genutzt mit  $T_1$  als Zeitpunkt des Wechsels und  $\tau$  Zeit nach dem Wechsel. Die Kosten-Faktoren werden festgelegt als  $\eta_y = \eta_z = 0.1$  und die anderen Parameter mit  $\alpha = 4, \beta = 10$  und  $j_{\text{out}} = 0.5$ . Die Fitness  $w$  zeigt über die gesamte Zeit keine Negativität.

Das Speicher-Modell aus Gleichung (8.17) und die nichtlinearen Produktionsfunktionen aus Gleichungen (8.8) und (8.9) ergeben das nichtlineare Speichermodell.

Betrachtet man das Wachstum  $w_k$  des nichtlinearen Modells mit Speicher, so kann man erkennen, dass genau wie bei dem linearen Modell mit Speicher auch hier durch die Einführung des Energiespeichers das Modell kein

negatives Wachstum produziert und somit kein potentiell Absterben induziert wird (siehe Abbildung 8.10). Direkt nach dem ersten Wechsel des Signals bei  $t = T_1 + \tau$ , mit  $T_1$  als Zeitpunkt des Wechsels und  $\tau$  als Zeit danach, zeigen die momentane Fitness  $w_k$  und das momentane Wachstum  $\dot{w}_k$  keine Negativität. Das Wachstum des nichtlinearen FFL-Modell mit einem Energiespeicher ist in Abbildung 8.10 berechnet mit Kosten-Faktoren  $\eta_y = \eta_z = 0.1$  dem Hill-Koeffizient  $h = 3$ , der Sättigungskonstante  $K = 0.1$  und den Parametern  $\alpha = 4$ ,  $\beta = 10$  und  $j_{\text{out}} = 0.5$ .

### Maximum der basalen Expression

Die Verwendung des FFL-Modells mit Speicher und nichtlinearen Produktionsfunktionen erzeugt ausschließlich positives Wachstum und positive Fitness und zeigt somit im Gegensatz zum Modell ohne Speicher in diesem Punkt einen evolutionären Vorteil. Das Modell mit Speicher soll die gleiche Optimalität in der basalen Expression  $k$  zeigen wie das Modell ohne Speicher. Dies wird in Lemma 4 gezeigt unter der Bedingung, dass die Produktionsfunktionen bei beiden Modellen gleich sind. Ist die Optimalität der basalen Expression  $k$  bei beiden Modellen gleich, so hat das nichtlineare Modell mit Speicher ebenfalls eine streng optimale basale Expression.

**Lemma 4.** *Unter der Voraussetzung, dass  $s_x$  periodisch ist, die Produktionsfunktionen gleich sind und das System  $(x, y_k, z_k, a_k, b_k)$  somit ebenfalls zu einer periodischen Lösung tendiert, optimieren die gleichen Parameter die Nutzenbilanz-Funktion des Modells ohne Speicher wie die Fitness des Modells mit Speicher.*

**Beweis.** Es wird von

$$\lim_{T \rightarrow \infty} w_k(T)/T$$

und  $C(k)$  im Falle von periodischen Signalen und somit auch periodischen Lösungen aus mit einer Periode von  $T$ , bis auf  $w_k$  im Modell mit Speicher, ausgegangen. Werden die Differentialgleichungen über die erste Periode integriert, so kann für  $b_k$  aus Gleichungen (8.18) ( $\dot{b}_k = j_1 - j_2$ ) beschrieben werden und somit gilt

$$0 = \int_0^T \dot{b}_k dt = \int_0^{T_2} j_1 - j_2 dt.$$

Das Integral ist durch das periodische Signal gleich Null. Durch die Gleichung

$$\dot{a}_k = j_{\text{in}} - j_1 + j_2 - j_y - j_z - j_{\text{out}} - j_w$$

kann berechnet werden, dass

$$w_k(T) - w_k(0) = \int_0^T j_w dt = \int_0^T j_{\text{in}} - j_y - j_z - j_{\text{out}} dt$$

ist. Da für die Nutzenbilanz-Funktion

$$C(k) = \frac{1}{T} \int_0^T s_x (ky_0 + (1-k)L[y_0x]) - \eta_y (k + (1-k)x) - \eta_z x (k + (1-k)y_0) dt.$$

gilt und  $j_{\text{out}} = \text{const}$  und  $j_{\text{in}} = 1 - s_x + s_x z_k$ , gilt

$$\begin{aligned} w_k(T) - w_k(0) &= \int_0^T (1 - s_x) dt - j_{\text{out}} T + T C(k) \\ &= A + T C(k). \end{aligned}$$

$A$  ist nicht von den Modellparametern abhängig. Daher ist in jeder Periode die Fitness  $w_k$  gleich der Nutzenbilanz-Funktion  $C$  plus einer Konstante  $A$ .

Die Parameter, die  $C$  in dem Modell ohne Speicher optimieren, optimieren auch  $w_k$  in dem Speichermodell und umgekehrt.  $\square$

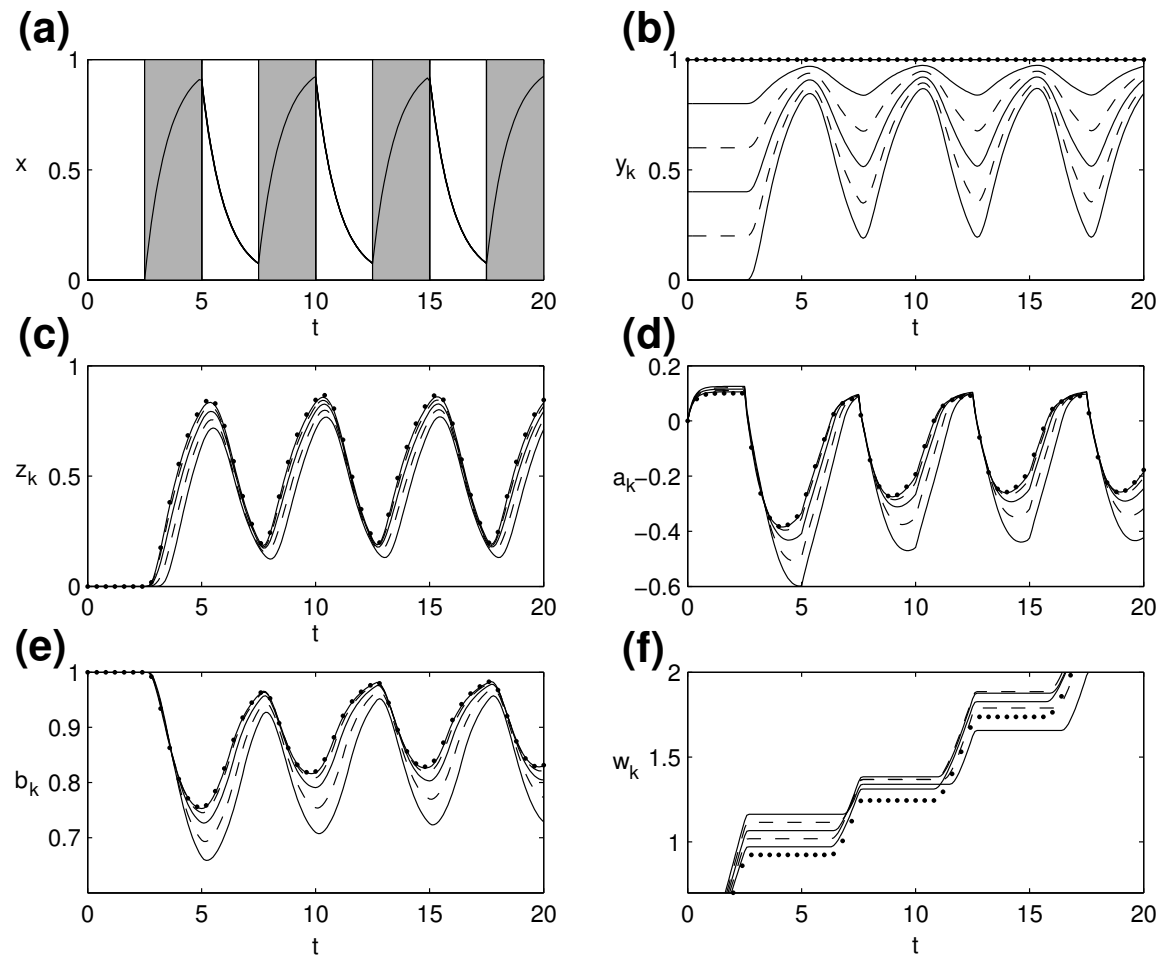
Die optimale basale Expression des Modells ohne Speicher unterscheidet sich nicht von der optimalen Expression des Modells mit Speicher bei gleichen Produktionsfunktionen. Dies kommt daher, weil der Energiespeicher keinen Einfluss auf die Optimalität der basalen Expression hat (Lemma 4). Somit konnte durch das Modell mit Speicher, die Negativität des Modells ohne Speicher eliminiert und das Verhalten des Modells in der Optimalität weiterhin beibehalten werden.

### Simulation mit nichtlineare Speichermodells

Das nichtlineare Speichersystem wird mit folgenden Anfangsbedingungen simuliert:  $x(0) = 0$ ,  $y_k(0) = k$ ,  $z_k(0) = 0$ ,  $a_k(0) = 0$ ,  $b_k(0) = b_{\text{set}} = 1$  und  $w_k(0) = 0$ . Das Signal hat den gleichen periodischen Aufbau wie in Kapitel 8.6. Die genutzten Parameter werden wie in den anderen Simulationen gewählt.

Die Fitness  $w_k$  des nichtlinearen Speichermodells hängt, genau wie das lineare Speichermodell, von den Wechseln des Signals ab. Der wichtigste Unterschied ist, dass die Fitness  $w_k$  in dem nichtlinearen Modell auch am größten für eine basale Expression ungleich null oder eins ( $k \in ]0, 1[$ ) sein kann. Dies geschieht, weil direkt nach einer Umschaltung zu Bereichen ohne Signal ( $s_x = 0$ ) die nichtlinearen Produktionsfunktionen das Wachstum ( $\dot{w}_k$ ) beeinflussen. Man kann den Unterschied der beiden Modelle in Abbildung 8.14 erkennen.



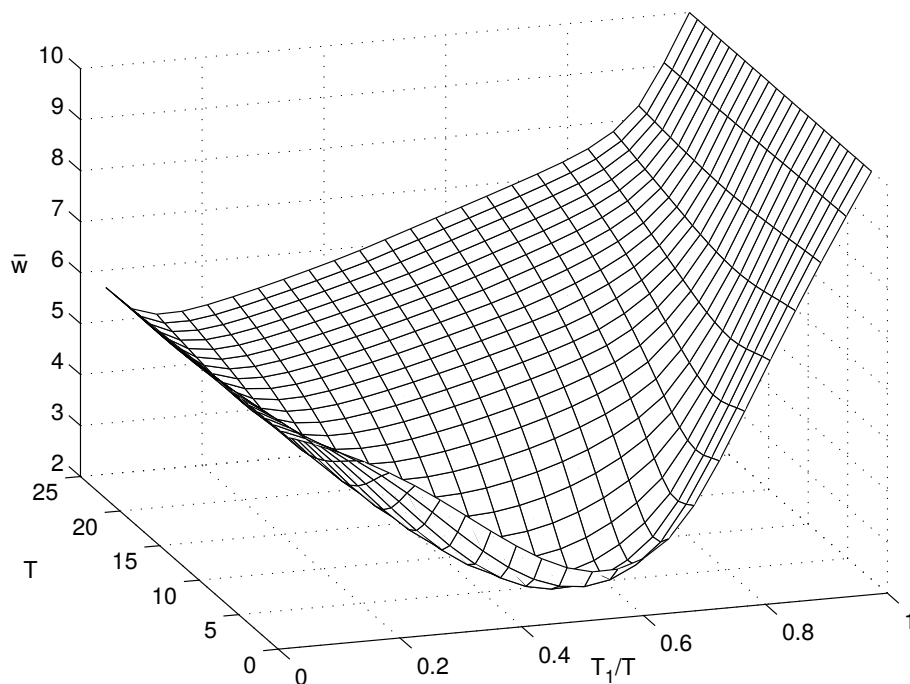


**Abbildung 8.11.: Simulationsergebnis des nichtlinearen Modells mit Speicher**

Simulation für  $k = 0, 0.2, 0.4, 0.6, 0.8, 1$  (durchgängige, gestrichelte, durchgängige, gestrichelte, durchgängige und gepunktet Linien). Der erste Regulator  $x$  und die Phasen ohne Signal werden mit grauem Hintergrund in (a) gezeigt. (b) zeigt den Verlauf des zweiten Regulators  $y_k$ . Die Zielvariable  $z_k$  befindet sich in Abbildung (c). (d) und (e) zeigen je den Verteiler- und den Speicherverlauf, welche abhängig von  $x$ ,  $y_k$  und  $z_k$  steigen und fallen. Die Fitness  $w_k$  steigt monoton (f). Wenn in dem Verteiler  $a_k$  Energie vorliegt, kommt es zu Wachstumsschüben. Zu allen anderen Zeiten ist die Fitness konstant.

## Fitness

Simuliert man das nichtlineare Modell mit Speicher mit verschiedenen Periodenlängen  $T$ , hier  $T \in ]0, 25]$ , und unterschiedlichem Anteil ohne Signal  $T_1/T$  mit  $s_x = 0$  für  $t \in ]0, T_1[$ , so kann für jede dieser periodischen Umweltbedingungen die optimale Fitness  $w_k$  zu jeder basalen Expression  $k$  berechnet werden. Die optimale Fitness wird derart berechnet, dass zunächst genügend viele Perioden (hier 15) als Einschwingphase genutzt werden. Die Fitness wird über darauffolgende 20 Perioden integriert und über die Zeit normiert.



**Abbildung 8.12.: Optimale Fitness des nichtlinearen Speichersmodells**

Die optimale normierte Fitness  $\bar{w}$  des nichtlinearen Modells mit Speicher mit Periode  $T$  und dem Anteil mit  $s_x = 0$  als  $T_1/T$

Die optimale normierte Fitness

$$\bar{w} = \max_k ((\int_0^T w_k(k) dt)/(T))$$

wird numerisch berechnet und in Abbildung 8.12 gezeigt. Genutzt werden die gleichen Parameter wie in der Simulation.

In Abbildung 8.12 wird gezeigt, dass das nichtlineare Speichermodell die größte durchschnittliche Fitness  $\bar{w}$  für besonders lange Phasen ohne Signal,  $T_1/T$  mit  $s_x = 0$  für  $t \in ]0, T_1[$  besitzt. Gerade bei einem stetigen Wechsel zwischen Phasen mit und ohne Signal und kurzen Gesamtperioden entsteht eine schlechtere Fitnessbilanz, hervorgerufen durch einen schnellen gleichmäßigen Wechsel zwischen den Signalen. Lebt ein Bakterium in einer solchen Umgebung, ist es besonders wichtig, dass die Fitness optimiert wird, da es sonst zu gar keinem Wachstum mehr kommen kann und somit die Bakterien möglicherweise sterben könnten.

### **Optimale basale Expression des nichtlinearen Modells mit Speicher**

Die optimale basale Expression  $k$  des nichtlinearen Modells mit Speicher wird zu der maximalen Fitness  $w_k$  (siehe für das nichtlineare Modell Abbildung 8.12) berechnet. Es wird ein periodisches Signal mit der maximalen Periodenlänge von  $T = 40$  genutzt. Zu jeder Periodenlänge von  $T \in ]0, 40]$  und Anteil der Zeit ohne Signal ( $T_1/T \in [0, 1]$ ) wird das optimale  $k$  berechnet.

In Abbildung 8.13 ist die optimale basale Expression des nichtlinearen Speichermodells dargestellt. Mittels Farbcode kann man erkennen, welche basale Expression zu welcher Periode optimal ist. Es werden nur periodische Signale mit einer maximalen Länge von  $T = 40$  betrachtet. Dunkelblaue

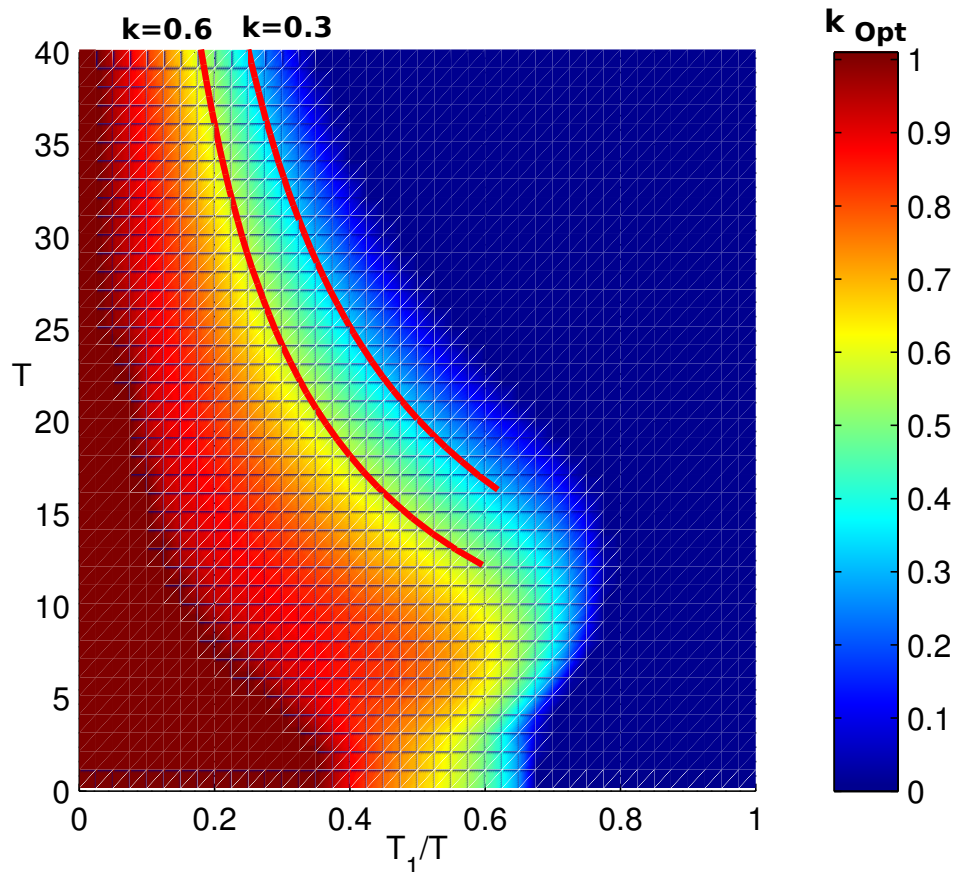
Stellen zeigen, dass hier ein maximales Wachstum für fehlende basale Expression vorliegt ( $k = 0$ ). Bei dunkelroter Farbe ist das optimale Wachstum für maximale basale Expression vorhanden ( $k = 1$ ). Alle Farben zwischen rot und blau zeigen ein optimales Wachstum für eine basale Expression zwischen null und eins ( $k \in ]0, 1[$ ).

Man kann deutlich erkennen, dass abhängig von dem Signal, eine basale Expression zwischen null und eins optimal sein kann, also ein streng optimales  $k$  vorliegt. Im nichtlinearen Modell kommt es zu einem streng optimalen  $k$ , im linearen Modell gibt es kein streng optimales  $k$ . Eine notwendige Bedingung für eine streng maximale basale Expression sind also die nichtlinearen Produktionsfunktionen.

Innerhalb der Zeitspanne ohne Signal ( $t \in [0, T_1]$ ) entstehen nur energetische Kosten durch die basale Expression. Der positive Nutzen der basalen Expression in dem System entsteht nur kurz nach dem Wechsel zum Signal mit Stimulus ( $t \in \{T_1, T\}$ ). Dort führt eine höhere basale Expression zu einem schnelleren Wachstum der Zielvariable und somit zu einem schneller einsetzenden Nutzen. Ohne die basale Expression oder mit niedriger basaler Expression wird mehr Zeit benötigt. Bei Erreichen des Gleichgewichtszustandes der Variablen existiert kein Unterschied mehr zwischen den basalen Expressionen. Der Nutzen unterscheidet sich je nach basaler Expression dementsprechend nur direkt nach einem Wechsel des Signals.

Sind die Zustandsvariablen bei Phasen mit Signal ( $t \in \{T_1, T\}$ ) auf dem Gleichgewichtszustand, also wenn die Gesamtperiode ( $T$ ) lang genug ist, ist der aerobe Anteil immer gleich lang für gleich große basale Expression  $k$ . Dies kann man gut an den roten Linien in Abbildung 8.13 erkennen. Die roten Linien zeigen das optimale Wachstum für  $k = 0.3$  bzw.  $k = 0.6$ . Für die basale Expression von  $k = 0.3$  ist die Länge des Signals ohne Stimulus (also  $T_1$ ) immer bei ca. 10 Zeiteinheiten, bei  $k = 0.6$  bei ca. 7.2 Zeiteinheiten. Dieses ist

unabhängig von der Länge der Phase mit Signal ( $T - T_1$ ) oder der Gesamtlänge der Periode ( $T$ ), gilt jedoch nur ab einer gewissen Gesamtperiodenlänge (hier erst bei ca.  $T > 12$ ).



**Abbildung 8.13.: Optimale basale Expression des nichtlinearen Speichermodells**

Optimale basale Expression ( $k$ ) des nichtlinearen Modells mit Speicher, mit  $T$  als Gesamtperiodenlänge und  $T_1/T$  als Anteil ohne Signal. Dunkelblau zeigt ein optimales Wachstum  $w_k$  für keine basale Expression ( $k = 0$ ). Bei dunkelrot liegt ein maximales Wachstum bei voller basaler Expression ( $k = 1$ ). Die Farben von dunkelblau über gelb bis dunkelrot stellen eine basale Expression zwischen null und eins dar ( $k \in ]0, 1[$ ) vor. Die roten Linien stellen jeweils die optimale basale Expression für lange Perioden für  $k = 0.6$  bzw.  $k = 0.3$  dar.

Die Abbauraten der Zustandvariablen sind in diesem Modell unabhängig

von der basalen Expression. In Phasen ohne Signal werden die Zustandsvariablen abgebaut. Alle Variablen müssen hinreichend abgebaut werden, um den vollen Nutzen aus dem darauffolgenden Wechsel zu ziehen, so dass ein Unterschied zwischen den verschiedenen basalen Expressionen des zweiten Regulators deutlich wird. Nur wenn der Regulator so weit abgebaut wird, dass bei einem erneuten Wechsel zur Signalanwesenheit die Zeitverzögerungsunterschiede der Produktion der Zielgene je nach basaler Expression groß genug sind, kann sich dort der notwendige Nutzen ergeben. Die basale Expression verursacht in dieser Zeit Kosten. Direkt nach Aktivierung des Signals bauen sich die Zustandsvariablen wieder auf. Hierbei ist das System mit basaler Expression besser vorbereitet, wenn ein Wechsel des Signals stattfinden sollte, die Verzögerung ist kürzer und aus der Zielvariable kann schneller Nutzen gezogen werden. Mit größer werdender basaler Expression werden auch die energetischen Kosten größer. Also ist eine ausgewogene basale Expression für diese Signalwechsel von Vorteil. Bei längeren Phasen ohne Signal verursacht die basale Expression zu viele Kosten, um diese durch den Nutzen der verkürzten Verzögerung ausgleichen zu können.

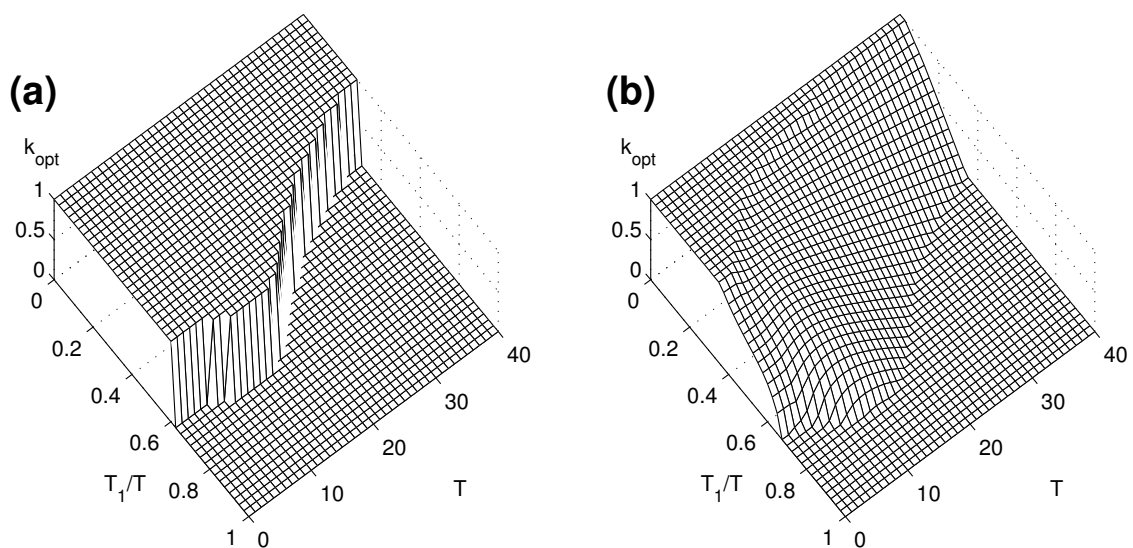
Bei kürzeren Gesamtperiodenlängen  $T$  ( $T < 12$ ) können die Zustandsvariablen nicht ihren Gleichgewichtszustand erreichen, so dass der Vorteil der verkürzten Verzögerung geringer ist. Nur kleine Kosten für die basale Expression können durch die verkürzte Verzögerung der Produktion ausgeglichen werden. Wenn nur ein geringer Vorteil entsteht, so können auch nur kürzere Phasen mit Signal ausgeglichen werden und somit von Vorteil sein. Dieses lässt sich in Abbildung 8.13 durch den Vorsprung erkennen.

Kosten entstehen bei Phasen ohne Signal. Nutzen entstehen nach dem Wechsel von Phasen ohne Signal zu Phasen mit Signal. Es ist notwendig, dass ein stetiger Wechsel zwischen beiden Zuständen vorliegt, um dieses System als

optimal zu bestimmen. Das Modell zeigt gut, dass genau eine Mischung aus Signal-Anwesenheit und Abwesenheit zu einem ausgewogenen Verhältnis zwischen Nutzen und Kosten der basalen Expression kommt.

### 8.3.5. Vergleich des linearen und des nichtlinearen Modells mit Speicher

In Abbildung 8.14 werden das lineare und das nichtlineare Modell mit Speicher innerhalb eines 3-dimensionalen Koordinatensystems dargestellt und verglichen. Wie in Abbildungen 8.9 und 8.13 wird die optimale basale Expression des nichtlinearen Speichersystem in Abbildung 8.14 (a) und (b) dargestellt.



**Abbildung 8.14.: Optimal basale Expression des linearen und nichtlinearen Speichersmodells**

Die Abbildung zeigt ein optimales Wachstum für eine basale Expression zwischen null und eins ( $k \in ]0, 1[$ ). (a) Optimale basale Expression ( $k$ ) des linearen Modells mit Speicher, mit  $T$  als Gesamtperiodenlänge und  $T_1/T$  als Anteil des Parts mit  $s_x = 0$ . (b) Optimale basale Expression ( $k$ ) des nichtlinearen Modells mit Speicher, mit  $T$  als Gesamtperiodenlänge und  $T_1/T$  als Anteil des Parts mit  $s_x = 0$ .

Man kann die optimale basale Expression  $k_{\text{opt}}$  in Abbildung 8.14 (a) des linearen und in (b) des nichtlinearen Speichermodell erkennen. In Abbildung 8.14 (a) wird eine Stufe zwischen keiner und voller basaler Expression dargestellt. Die optimale basale Expression  $k$  ist entweder null oder eins, also  $k \in \{0, 1\}$ . Ein strenges Maximum existiert in dem linearen Modell für die basale Expression nicht. Ob keine oder eine volle basale Expression optimal ist, hängt von dem Signal ab.

Das Ergebniss des nichtlinearen Modell mit Speicher wird in Abbildung 8.14 (b) die optimale basale Expression  $k$  gezeigt. Hier gibt es einen kontinuierlichen Übergang zwischen einem optimalen  $k = 0$  und einem optimalen  $k = 1$ . Es existiert für dieses nichtlineare Modell mit Speicher eine streng maximale basale Expression für bestimmte Signallängen. Um eine streng maximale Expression zu erreichen, sind folglich die nichtlinearen Produktionsfunktionen eine notwendige Bedingung.

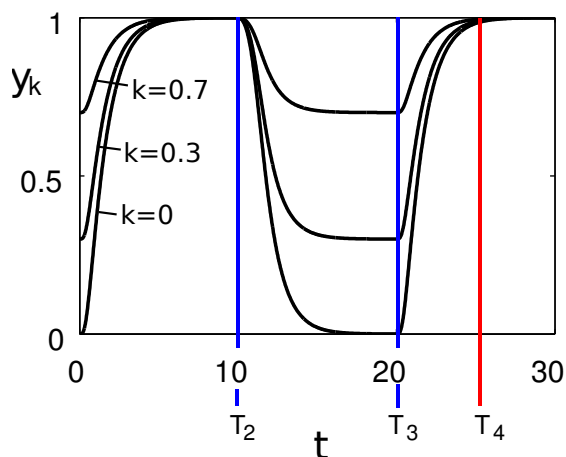
### 8.3.6. Aufbau des zweiten Regulators nach Signalwechsel

Wenn nach der Phase mit Signal  $s_x = 1$  die Konzentration des zweiten Regulators  $y_k$  mit großer basaler Expression ebenfalls größer ist, kann es ein, dass es zu einem schnelleren Aufbau derselbigen kommt, wenn das Signal zurück wechselt auf  $s_x = 1$ . Mittels Lemma 3 kann dieses gezeigt werden. Genutzt werden die gleichen Parameter wie in den vorangegangenen Simulationen. Das Signal wird definiert als:



$$\bar{s}_x = \begin{cases} 0 & \text{wenn } t \in [0, T_1[ \\ 1 & \text{wenn } t \in [T_1, T_2[ \\ 0 & \text{wenn } t \in [T_2, T_3[ \\ 1 & \text{wenn } t \in [T_3, T[. \end{cases}$$

$T_2$  wird derart gewählt, dass sich die Zustandsvariablen im Gleichgewichtszustand des Maximums befinden ( $\max(x) = \max(y_k) = \max(z_k) = 1$ ). Diese Länge der des Signals  $\bar{s}_x = 1$  auf dem Intervall  $[T_2, T_3[$  wird variiert, um herauszufinden, wie lang ein Gedächtniseffekt vorhanden ist. Als Gedächtnis wird bezeichnet, ob eine Variable in dieser Zeit noch nicht vollständig abgebaut wurden und somit schneller wieder aufgebaut werden können (siehe Definition 10).

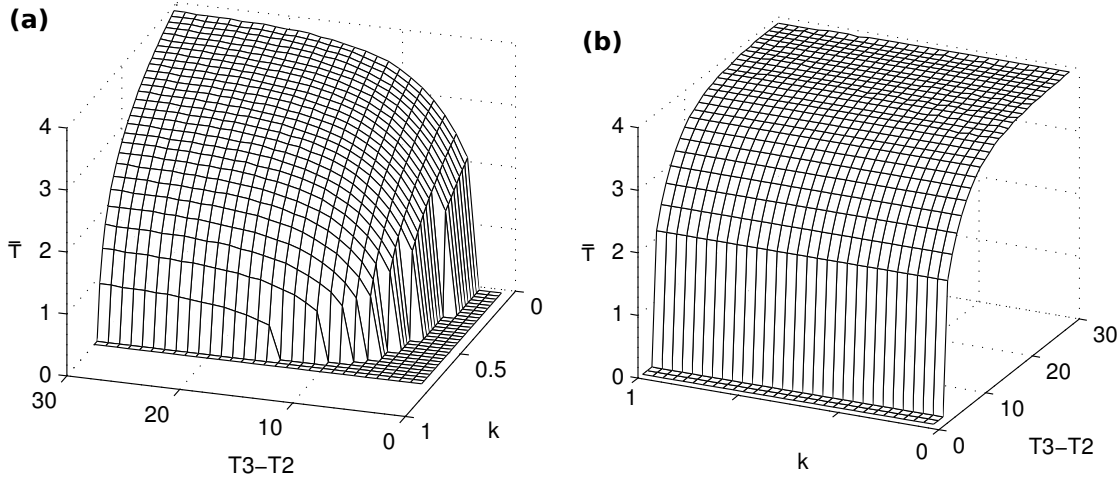


### Abbildung 8.15.: Abbau und erneuter Aufbau des zweiten Regulators

Für drei verschiedene basale Expressionen ( $k \in \{0, 0.3, 0.7\}$ ) wird der Aufbau und der erneute Abbau gezeigt. Je nach basaler Expression ist der erneute Aufbau unterschiedlich. Hier ist  $T_2 = 10$  und  $T_3 = 20$ .

Der zweite Regulator  $y_k$  wird ab dem Zeitpunkt  $T_3$  wieder aufgebaut. Nach erneutem Anschalten des Signals wird der zweite Regulator  $y_k$  wieder aufge-

baut. Erreicht dieser 95% seiner Gleichgewichtskonzentration ist der Zeitpunkt  $T_4$  erreicht. Somit ist  $T_4 \in [T_3, T]$ . Dieses wird in Abbildung 8.15 verdeutlicht.



**Abbildung 8.16.: Benötigte Zeit zum Aufbau des zweiten Regulators nicht-normiert und normiert.**

Die benötigte Zeit, um den zweiten Regulator  $y_k$  wieder aufzubauen und 95 % des Gleichgewichtszustandes zu erreichen, wird mit  $\bar{T}$  beschrieben und numerisch berechnet.  $T_3 - T_2$  ist die Zeit der Signal-Anwesenheit und  $k$  die unterschiedlichen basalen Expressionen. (a) zeigt die benötigte Zeit für den Aufbau des nicht-normierten zweiten Regulator  $y_k$  und (b) zeigt die benötigte Zeit für den Aufbau des normierten zweiten Regulator  $y_k$ .

Die benötigte Zeit  $\bar{T} = T_4 - T_3$  zum erneuten Aufbau wird numerisch für alle basalen Expressionen und unterschiedliche Längen der Signalabwesenheit berechnet und in Abbildung 8.16 dargestellt. Für den Fall, dass  $T_3 - T_2$  besonders klein ist, kann man erkennen, dass die benötigte Zeit für den Aufbau  $\bar{T}$  besonders gering ist.

Der Aufbau lässt sich nach der basalen Expression normieren. Es wird nach numerischer Berechnung des zweiten Regulators  $y_k$  diese Größe auf  $\bar{y}_k = \frac{y_k - k}{1 - k}$  normiert. Danach wird ebenso wie in Abbildung 8.16 (a) die benötigte Zeit berechnet, um den zweiten Regulator wieder aufzubauen. Man kann in Abbildung 8.16 (b) nun gut erkennen, dass die Zeit für einen erneuten Aufbau nicht mehr von der basalen Expression abhängt.

## 8.4. Evolutionärer Vorteil der Modelle

Als evolutionär vorteilhaft wurde ein System definiert, welches ein positives Wachstum und eine streng maximale Fitness, bzw. Nutzenbilanz, hat (siehe Definition 9). Beide Bedingungen müssen für ein Modell gelten, so dass es den Vorteil des Systems der FFL mit basaler Expression darstellen kann.

Das erste gezeigte Modell der Modellfamilie war ein möglichst einfaches System mit linearen Produktionsfunktionen und ohne Energiespeicher. Das lineare Modell konnte keine der Bedingungen des evolutionären Vorteils zeigen. Durch Erweiterung des Modells durch nichtlineare Produktionsfunktionen (siehe Kapitel 8.2) konnte gezeigt werden, dass eine Nutzenbilanz streng maximal in Abhängigkeit von der basalen Expression unter bestimmten Voraussetzungen sein kann. Hierbei ist die verkürzte Zeitverzögerung bei der Produktion der Zielvariable ein größerer Vorteil als die verbrauchte Energie durch die Produktion der basalen Expression. In diesem Modell wird jedoch keine Energie gespeichert. Wird Energie benötigt, ohne dass gleichzeitig Energie produziert wird, kommt es zu einem Sinken der Nutzenbilanz. Die verursachten Kosten nach einem Signalwechsel können innerhalb des Modells ohne Speicher jedoch nicht abgefangen werden, welches zu nicht wachsenden oder sogar möglicherweise sterbenden Bakterien führen kann. Dies ist nicht evolutionär vorteilhaft.

Die notwendige Erweiterung des Modells erfolgte durch einen Energiespeicher (siehe Kapitel 8.3), auf den jedes Bakterium zurückgreifen kann. Beinhaltet der Speicher genügend Energie, kann diese Energie in das Wachstum gesteckt werden. Mittels des eingeführten Energiespeichers kommt es nicht mehr zu einem Sinken der Fitness, also nicht zu einem negativen Wachstum.

Die von den Modellen ohne Speicher nicht erfüllte Bedingung des evolutionären Vorteils wurde somit vom Modell mit Speicher erfüllt. Es zeigt sich, dass auch hier nur mittels nichtlinearer Produktionsfunktionen des zweiten Regulators und der Zielvariable eine streng maximale Fitness zwischen keiner und voller basalen Expression des zweiten Regulators existiert. Durch den eingefügten Energie-Speicher in das Modell kann gezeigt werden, dass eine FFL mit basaler Expression evolutionär vorteilhaft ist. Es ist notwendig, um die beide Bedingung des evolutionären Vorteils zu erfüllen, dass sowohl ein Energiespeicher, welcher innerhalb der Bakterien zum Beispiel durch ATP, PHA oder Polyphosphate existiert, als auch nichtlineare Produktionsfunktionen in dem Modell eingeführt werden.

Je nach Umgebung kann die Fitness unterschiedlich groß sein, auch die optimierte Fitness. In Abbildung 8.12 wird zu jeder Umgebung die optimale Fitness abhängig von  $k$  berechnet und nach der Zeit und den Perioden normiert. Eine schnell wechselnde Umgebung kostet dem Bakterium viel Energie und es ist daher besonders wichtig, dass die basale Expression optimal ist. In Umgebungen, in denen eine basale Expression nicht vorteilhaft erscheint, ist die Fitness an sich bereits sehr groß, somit bringt eine basale Expression keinen Überlebensvorteil. Nur bei einem stetigen Wechsel der Umwelt ist es evolutionär vorteilhaft, eine basale Expression zu nutzen. Die streng optimale Fitness mit der dazugehörigen basalen Expression  $k$  hängt daher direkt vom Wechsel der Umgebung, also von dem Signal  $s_x$ , ab. Es wurde gezeigt, dass das Modell mit basaler Expression besser an wechselnde Umgebungen angepasst ist. Eine basale Expression ist hier evolutionär vorteilhaft.

## 8.5. *Pseudomonas aeruginosa*

Bei dem Beispiel der Regulation der Nitratatmung in *P. aeruginosa* stellen die Phasen ohne Signal, bei  $s_x = 0$ , die aeroben Phasen dar. Bei  $s_x = 1$  befindet sich das System in einer anaeroben Umgebung. Der zusätzliche Nutzen durch die basale Expression entsteht nur nach einem Wechsel von einer aeroben zu einer anaeroben Phase. Es ist notwendig für *P. aeruginosa* eine optimale Nutzenbilanz, um das Überleben zu sichern. Dieses Optimierungsproblem hängt direkt von der Umgebung und der basalen Expression  $k$  ab. Die Umgebung kann nicht von dem Bakterium beeinflusst werden.

In der schnell wechselnden Umgebung, in der eine basale Expression zwischen null und eins optimal erscheint, also  $k \in ]0, 1[$  (siehe Abbildung 8.13), ist ein Wachstum eher gering (siehe Abbildung 8.12). Es scheint für *P. aeruginosa* unumgänglich zu sein eine basale Expression des zweiten Regulators, dem NarXL-System, zu haben, um überleben zu können. Das Denitrifikationssystem von *P. aeruginosa* hat sich mittels einer basalen Expression von ca.  $k = 0.5$  optimal an einen ständigen Wechsel der Umgebung angepasst (Schreiber et al., 2007).

Da *P. aeruginosa* an eine wechselnde Umgebung angepasst ist, scheint dieses Bakterium ein typisches freilebendes Bakterium zu sein. Innerhalb der freien Umwelt kommt es im Gegensatz zu einer Wirtsumwelt zu ständigen Änderungen der Umgebungsbedingungen. Wäre *P. aeruginosa* ein typisches pathogenes Bakterium, wäre es möglicherweise nicht an wechselnde Umgebungen angepasst. Eine Ausnahme von einer Wirtsumwelt, bei der es ebenfalls zu häufigen Wechseln der Umwelt kommt, ist wahrscheinlich die Lunge von zystische Fibrose-Patienten, da diese durch Husten und Bewegungen des Patienten ständigen Erschütterungen ausgesetzt ist und somit aerobe wie anaerobe Teile des Mukus (siehe Kapitel 6.1) miteinander

mischt.

Die besondere Gefahr von *P. aeruginosa* ist unter anderem die Fähigkeit, Biofilme zu bilden, wie z. B. auch in der Lunge von zystische Fibrose Patienten. Innerhalb dieser Biofilme herrschen in den tieferen Schichten anaerobe Bedingungen. An der Oberfläche ist jedoch eine eher aerobe Umgebung (Xu et al., 1998). Wird der Biofilm durchmischt, kann es somit für die Bakterien auch zu einer Änderung der Umgebung (Lieleg et al., 2011) kommen.

Aus dieser Betrachtung könnte die Variabilität der Nitratatmung von *P. aeruginosa* ein Zeichen für Pathogenität sein. Es ist aber möglich, dass *P. aeruginosa* ursprünglich ein Umweltkeim war und sich durch die Mechanismen der Nitratatmungsregelung gut an eine Wirtsumwelt angepasst hat, somit die Pathogenität aus der Adaptionmöglichkeit und der metabolischen Vielseitigkeit des Bakteriums entstanden ist.

## 8.6. Ausblick

Die Ergebnisse der Modellierung der FFL mit basaler Expression wurden für das Regulationsmodell innerhalb der Nitratatmung in *P. aeruginosa* adaptiert, wobei die Parameter zufällig bestimmt wurden. Um die in dieser Arbeit getroffenen Aussagen für *P. aeruginosa* zu bestätigen, müssen die Parameter von der Regulation innerhalb der Nitratatmung bestimmt werden.

Durch die theoretische Modellierung dieses Regulationsmotiv können die Ergebnisse dieser Arbeit ebenfalls auf andere Bakterien mit der gleichen Regulation angepasst und angewandt werden, wobei die Parameter dieser Regulationen ebenfalls bestimmt werden müssten. Die Ergebnisse können für die gleiche Regulation in anderen Bakterien adaptiert werden.

## **Teil III.**

# **Modellierung einer gepaarten positiven und negativen Feedback-Schleife von RovA in pathogenem *Yersinia* *pseudotuberculosis***





## 9. Einleitung

*Yersinia pseudotuberculosis* ist eine *Yersinia*-Art (so wie ebenfalls *Yersinia enterocolitica* und *Yersinia pestis*), die pathogen für den Menschen ist (Bottone, 1997, Dube et al., 2003). Der Hauptwirt dieses Bakteriums sind Nutztiere, aber der Mensch kann ebenfalls von *Y. pseudotuberculosis* befallen werden. Eine Infektion von *Y. pseudotuberculosis* wird durch RovA, einem globalem Transkriptionsfaktor und Virulenzregulator, gesteuert.

RovA wird durch zwei Feedback-Schleifen (FBL, engl. feedback loop), einer positiven und einer negativen, reguliert (siehe Kapitel 2.6.1). Beide Feedback-Schleifen, ebenso wie der Abbau von RovA werden durch die Temperatur beeinflusst. Die RovA-Konzentration zeigt eine temperaturabhängige Bistabilität und eine Hysterese. Dieses Phänomen ist von entscheidender Bedeutung für das Bakterium bei der Unterscheidung zwischen Wirts- und Umgebungstemperatur. Innerhalb dieser Arbeit wird eine mathematische Differentialgleichung erstellt, um die Regulation von RovA in *Y. pseudotuberculosis* zu beschreiben und zu analysieren. Es werden zunächst alle Parameter durch gegebenen Beobachtungen berechnet und dann wird die Temperaturgenauigkeit, die Bistabilität in den Gleichgewichtszuständen und der Einfluss der einzelnen berechneten Parameter des Modells analysiert.

## 9.1. *Yersinia pseudotuberculosis*

Das enteropathogene Bakterium *Y. pseudotuberculosis* ist ein aerobes Gram-negatives Bakterium mit einer Größe von ca.  $1.5 - 6\mu\text{m} \times 0.4 - 1\mu\text{m}$ . Das Bakterium besitzt eine Geißel und ist bei Zimmertemperatur (ca.  $20^\circ\text{C}$ ) mobil. *Y. pseudotuberculosis* ist resistent gegen viele äußere Umwelteinflüsse. Es kann lange Zeit im Boden, im Wasser aber auch im Blut von Wirten überleben und bleibt dabei weiterhin infektiös (Heroven et al., 2012). Eine Infektion mit *Y. pseudotuberculosis* wird als Yersiniose bezeichnet (Tran et al., 2005). Zumeist infiziert *Y. pseudotuberculosis* Tiere wie zum Beispiel Nager, Schweine oder Wildvögel, ebenso Menschen. Hierbei kommt es zu Tuberkolose-ähnlichen Symptomen, wie lokale Gewebenekrosen oder Granulome (Entzündungsbedingte, knotenartige Gewebeneubildung) in Lymphknoten oder Leber.

### 9.1.1. Infektion

Innerhalb des Menschen ist eine Infektion mit *Y. pseudotuberculosis* dem Scharlach-Fieber sehr ähnlich, auch Izumi-Fieber genannt. Infizierte Tiere (auch Haustiere) fungieren als Überträger, aber auch über Nahrungsmittel oder Wasser können die Krankheitserreger übertragen werden. Die Inkubationszeit beträgt meistens fünf bis zehn Tage, maximal drei Wochen. Die Infektion führt zu Fieber, Dünndarmerkrankungen, Durchfall und einem Befall der Lymphknoten (Koornhof et al., 1999, Bottone, 1999). Besonders Kinder und junge Erwachsene sind anfällig für eine *Y. pseudotuberculosis*-Infektion, die meist oral übertragen werden (Heroven et al., 2012). Sowohl Impfungen als auch Antibiotika werden gegen *Y. pseudotuberculosis* eingesetzt.

Normalerweise liegt die Umgebungstemperatur dieser Bakterien unter  $30^\circ\text{C}$ .

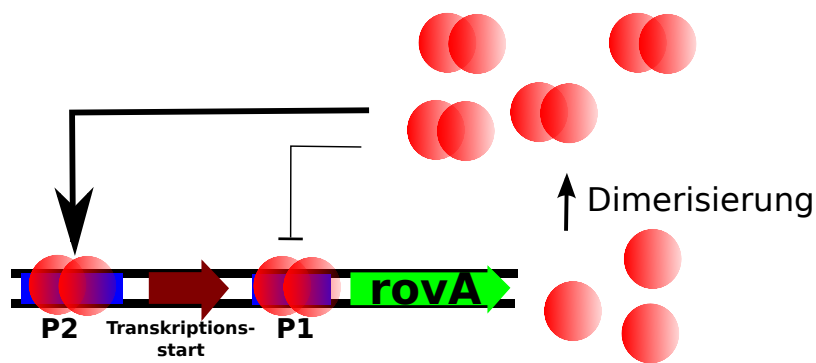
Erhöht sich die Temperatur auf über 37°C, die einer Wirtstemperatur, so wird dies von *Y. pseudotuberculosis* als Signal gewertet, um ihr Virulenz-Programm zu aktivieren (Konkel and Tilly, 2000). Es ist notwendig für pathogene Bakterien, sich dieser Wirts-Umwelt schnell anzupassen. Bei allen Bakterien kommt es bei einer Anpassung an eine veränderte Umwelt zu einer Reihe von Aktivierungen oder Reprimierungen von Enzymen (Quade et al., 2012). Die meisten Pathogenitätsfaktoren von Yersinien werden thermisch geregelt. Die Lon-Protease baut RovA bei höheren Temperatur (ab ca. 37°C) verstärkt ab (Herbst et al., 2009). Die Autoregulation von RovA ist ebenfalls abhängig von der Temperatur (Quade et al., 2012).

Das äußere Membranprotein Invasin unterstützt effizient die Aufnahme der Bakterien in die M-Zellen (engl. microfold cells). M-Zellen sind spezialisierte Epithelzellen, die in der Wand des Dünndarms sitzen. Nachdem *Y. pseudotuberculosis* in den Dünndarm des Wirtes gelangt ist, zum Beispiel durch kontaminiertes Essen, muss das Virulenzenzym Invasin von dem Bakterium produziert werden, damit es in die M-Zellen in den Wirt eindringen kann. Es wird gezeigt, dass die Expression des Invasin-Gens (*inv*) abhängt von Temperatur, pH-Wert und der Wachstumsphase des Bakteriums (Dube et al., 2003). Diese Regulation der Invasin-Expression geschieht über den Transkriptionsregulator RovA (Tran et al., 2005). Invasin wird nur dann produziert, wenn das RovA-Level sehr gering ist (Revell and Miller, 2000, Nagel et al., 2001). Bei Anwesenheit von Invasin in *Y. pseudotuberculosis* kann das Bakterium innerhalb des Dünndarms den Wirt befallen.

### 9.1.2. RovA

Die Regulation des Transkriptionsfaktors RovA in *Y. pseudotuberculosis* und anderen Yersinien ist notwendig für das Bakterium, um seine Virulenz

zu kontrollieren (Tran et al., 2005). Diese Steuerung erfolgt über autoregulative RovA-Dimere, welches für die Expression des Invasin-Gens (Böhme et al., 2012) und anderen Virulenz- und Stressgene zuständig sind (siehe Kapitel 9.1.1). Invasin unterstützt die Aufnahme der Bakterien in den Wirt. Durch die schnelle Regulation von Invasin durch RovA ist eine effiziente Kolonisierung und Infizierung möglich.



#### Abbildung 9.1.: Operon von RovA bei 25°C

Rote Kugeln stellen RovA bei 25°C dar. Zwei Kugeln nebeneinander symbolisieren RovA-Dimere. RovA wird bei 25°C produziert und bindet als Dimer an zwei Promotoren ( $P_1$  und  $P_2$ ). In blau sind die Promotor-Bindestellen dargestellt. Der Promotor  $P_1$  liegt stromaufwärts des Transkriptionsstarts (dunkelroter Pfeil) und ist stark aktivierend. Der Promotor  $P_2$ , stromabwärts des Transkriptionsstarts liegend, ist schwach reprimierend. Als grüner Pfeil wird das RovA-Gen (*rovA*) symbolisiert. Die Bindung von RovA an die Promotoren stellen zwei autoregulative Feedback-Schleifen dar.

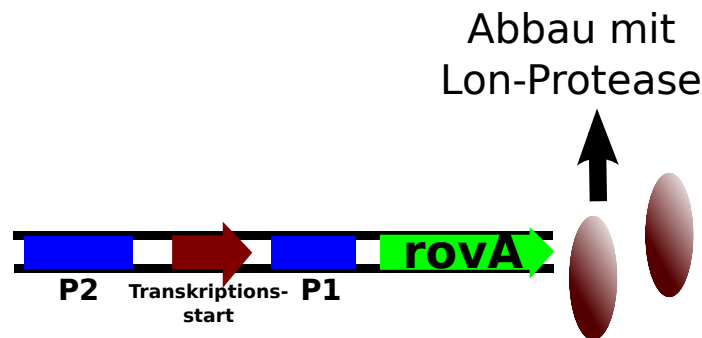
Das Operon von RovA bei einer Temperatur von 25°C wird in Abbildung 9.1 dargestellt. RovA als Dimer fungiert sowohl als Aktivator als auch als Repressor der eigenen Transkription. Es liegen daher zwei Autoregulationen vor. Unterhalb (stromabwärts) des Transkriptionsstarts liegt die Bindungsstelle des reprimierenden ( $P_1$ ) und stromaufwärts des aktivierenden Promotors ( $P_2$ ), (Tran et al., 2005). Beide Promotoren aktivieren, bzw. reprimieren das gleiche Gen. Sie stellen zwei autoregulatorische Feedback-Schleife (FBLs) dar. RovA wird somit durch das Produkt einer positiven und einer

negativen FBL reguliert (siehe Kapitel 2.6.1), (Heroven et al., 2004). An die Transkriptionsfaktorbindestellen innerhalb der Promotoren binden die RovA-Moleküle mit einer bestimmten Bindekonstante und aktivieren bzw. reprimieren die Expression somit. Im Folgenden wird die Bindekonstante an den aktivierenden Promotor mit  $k_a(T)$  bezeichnet und die Bindekonstante an den reprimierenden mit  $k_r(T)$ . Liegt RovA als Dimer vor, bindet es sowohl an den reprimierenden, als auch den aktivierenden Promotor innerhalb des RovA-Operons. Der reprimierende Promotor ist hierbei wesentlich schwächer als der aktivierende, somit kommt es zu einer stärkeren Expression von RovA.

Bei höheren Temperatur, ab ca. 37°C (Wirtstemperatur), ändert sich die Konformation von RovA, wodurch RovA nicht mehr als Dimer vorliegen kann und ein Abbau von RovA durch Proteasen begünstigt wird (siehe Kapitel 9.1.2). RovA wird von der Lon-Protease (Herbst et al., 2009) mit einer Rate von  $\delta(T)$  abgebaut. Die Bindung von RovA an beiden Promotoren ist bei höheren Temperaturen stark reduziert, Abbildung 9.2. Zusätzlich kann an die Promotoren von RovA bei höheren Temperaturen das Histon-ähnliche Protein H-NS binden (Tran et al., 2005). Ist H-NS sowohl an dem reprimierenden, als auch an dem aktivierenden Promotor gebunden, kann dann RovA nicht mehr binden. Es kommt insgesamt zu einer stärkeren Reprimierung des RovA-Operons als bei niedrigeren Temperaturen. Die RovA-Konzentration sinkt.

### **Temperaturabhängige Regulation von RovA**

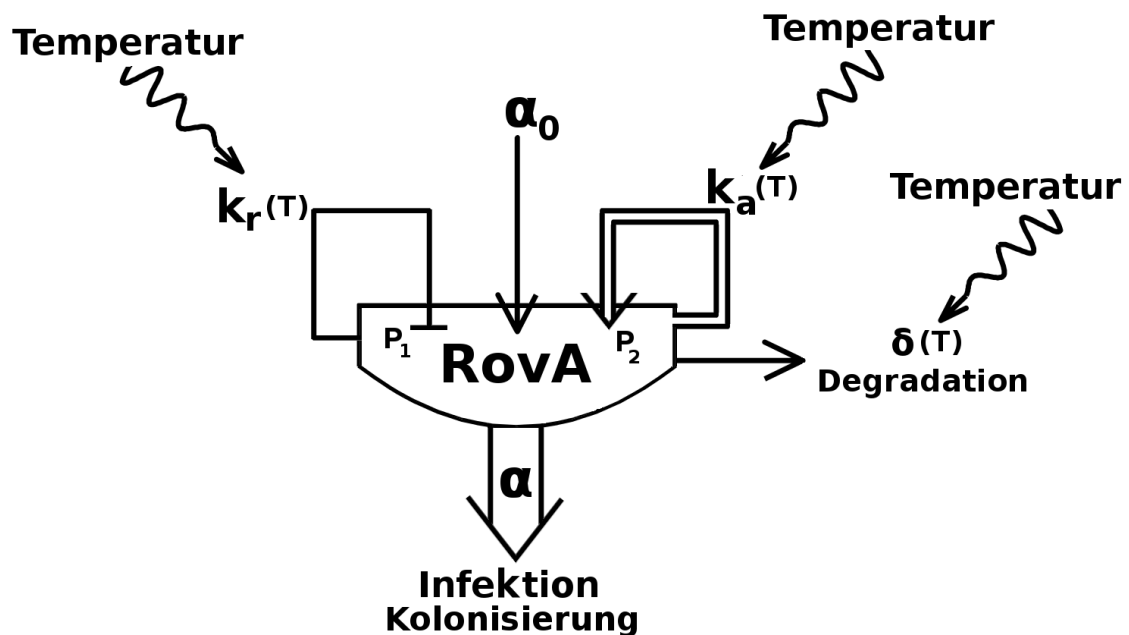
Die veränderte Produktion von RovA bei höherer Temperatur wird dadurch bedingt, dass RovA seine Konformation ändert (siehe Kapitel 9.1.2). Die Kristallisationsstruktur von RovA bei Temperaturerhöhung zeigt, dass sich die Dimerisationsdomäne von RovA bei höheren Temperaturen verändert.



### Abbildung 9.2.: Operon von RovA bei 37°C

Dunkelrote Ellipsen stellen RovA mit neuer Konformation dar, entstanden durch die höhere Temperatur. Durch neue Konformation bindet RovA nicht mehr an die Promotoren ( $P_1$  und  $P_2$  - blau). Die Lon-Protease kann RovA mit der neuen Konformation abbauen. Als dunkelgrüner Pfeil wird wieder das RovA-Gen (*rovA*) dargestellt.

Es kommt zu einer partiellen Entfaltung des Regulators, somit ist die Dimerisierung von RovA gestört (Quade et al., 2012). Dies hat zwei Auswirkungen auf die Regulation von RovA. Zum einem kann RovA dann als Autoregulator nur als Dimer an seine eigenen Promotoren  $P_1$  und  $P_2$  binden, zum anderen kann die Lon-Protease, aber auch andere Proteasen, nach der Konformationsänderung RovA vermehrt abbauen (Herbst et al., 2009). Die RovA-Produktion wird bei höheren Temperaturen ebenfalls durch das Histon-ähnliche Protein H-NS reprimiert, welches ab ca. 37°C an die Promotoren des RovA-Operons bindet und somit eine erneute Aktivierung verhindert (Tran et al., 2005) (siehe Kapitel 9.1.2). Durch die geringere Bindung von RovA an beide Promotoren, den erhöhten Abbau durch die Lon-Protease und der Bindung an die Promotoren von H-NS kommt es insgesamt zu einer geringeren Expression von RovA. Durch diese Einflüsse wird bei Temperaturen ab 37°C das *RovA* wesentlich geringer produziert (Herbst et al., 2009).



**Abbildung 9.3.: Regulationskreislauf von RovA abhängig von der Temperatur**

Die Regulation von RovA erfolgt über zwei Autoregulationen, einer positiven, 'starken' (gekennzeichnet durch den Doppelpfeil) und einer negativen 'schwachen' Regulation. Die Bindung an den aktivierenden Promotor (positive Autoregulation) erfolgt mit einer Bindekonstante von  $k_a(T)$ , die Bindung an den reprimierenden Promotor (negative Autoregulation) mit einer Bindekonstante von  $k_r(T)$ . Beide Bindekonstanten werden durch die Temperatur beeinflusst. Die basale Expression von RovA wird mit  $\alpha_0$  dargestellt und die maximale Expressionsrate mit  $\alpha$ . Beide sind temperaturunabhängig. Die Abbaurrate  $\delta(T)$  ist wiederum abhängig von der Temperatur.

In Abbildung 9.3 wird deutlich, auf welche Teile der Regulation von RovA die Temperatur einen Einfluss hat. Die Bindekonstante  $k_r(T)$  von RovA an den reprimierenden Promotor  $P_1$  wird ebenso durch die Temperatur beeinflusst, wie die Bindekonstante  $k_a(T)$  von RovA an den aktivierenden Promotor  $P_2$ . Beide Bindungen an die Promotoren stellen zwei Feedback-Schleifen dar. Da die Lon-Protease-Aktivität ebenfalls von der Temperatur abhängt, ist entsprechend auch die Degradation von RovA von der Temperatur abhängig. Ist RovA nur mit einer niedrigen Konzentration vorhanden, kann es

zur Infektion kommen. Die basale Expressionsrate von RovA wird mit  $\alpha_0$  beschrieben, die maximale Produktionsrate mit  $\alpha$ .

## 9.2. Feedback-Schleife

Feedback-Schleifen (FBLs) können sowohl in Form negativer als auch als positiver Rückkopplung auftreten (siehe Kapitel 2.6.1). Eine Besonderheit der RovA-Regulation ist, dass auf beide Arten von FBLs zurückgegriffen wird, die sich durch unterschiedliche Eigenschaften auszeichnen. Feedback-Schleifen sind häufig vorkommende Genregulationsmotive (Alon, 2007, Mitrophanov and Groisman, 2008).

### 9.2.1. Negative Feedback-Schleife

Die negative FBL (NFL) beschleunigt Prozesse (Oehler et al., 1990) und die benötigte Zeit, bis sich ein System im Gleichgewichtszustand befindet, verringert sich im Gegensatz zu Prozessen ohne NFL (Alon, 2007). Ist eine NFL in einem Prozess als Autoregulation, wie in RovA, involviert (siehe Kapitel 2.6.1), so wird die Produktion des Zielproteins von diesem selbst reprimiert. Eine NFL kann auch als ein Mehr-Gen-Motiv vorkommen.

Die NFL wird innerhalb dieser Arbeit mathematisch mittels Differentialgleichung dargestellt. Das Wachstum einer Zielvariablen wird mit einer NFL durch die Zielvariable  $x$  selbst reprimiert. Mittels der Hill-Funktion (siehe Kapitel 2.7) kann die Produktion der Zielvariable

$$\frac{dx(t)}{dt} = \alpha \frac{k^h}{k^h + x(t)^h}. \quad (9.1)$$



dargestellt werden, wobei  $(k^h)/(k^h + x(t)^h)$  der Anteil an Promotoren ist, an denen kein Repressor gebunden hat. Wird die Konzentration von  $x(t)$  immer größer, binden immer mehr Repressoren und der Anteil ohne Repressoren wird immer kleiner. Die Produktion des Proteins sinkt.

In Gleichung (9.1) ist  $k$  die Konstante, mit der die autoregulative Rückkopplung stattfindet. Um so größer diese Konstante ist, desto geringer ist der negative Einfluss der Zielvariable auf die Produktion. Innerhalb der Genregulation wäre die Rückkopplung die Bindung des Regulators an seinen eigenen Promotor. Mit einer maximalen Rate  $\alpha$  wird produziert. Der Parameter  $h$  stellt den Hill-Koeffizienten dar (siehe Kapitel 2.7). Steigt die Konzentration von  $x$ , so geht die Produktionsrate gegen Null.

Mit der Produktionskonstante  $\alpha$  wird die Zielvariable maximal produziert. Ist die Zielvariable nicht vorhanden, so wird diese maximal produziert, da keine Reprimierung statt findet. Mit steigender Konzentration der Zielvariable wird die Autoreprimierung stärker und dadurch sinkt die Produktion der Zielvariable. Eine NFL beschleunigt die Reaktionszeit. Durch die Autorepression kann es zu einer kleineren Gleichgewichtskonzentration kommen, dieses kann durch eine erhöhte Produktionskonstante  $\alpha$  ausgeglichen werden.

### 9.2.2. Positive Feedback-Schleife

Eine positive FBL (PFL) kann sowohl aus einem als auch mehreren Genen bestehen. So ist eine doppelte negative FBL ebenfalls eine PFL. Innerhalb der RovA-Steuerung existiert eine PFL als Autoregulation.

Eine PFL kann in biologischen Systemen ein bistabiles Verhalten auslösen (Hasty et al., 2002). Im Gegensatz zur NFL verlängert die PFL die benötigte

Zeit bis zum Gleichgewichtszustand. Ein klassisches Beispiel für eine PFL ist das *lac*-Operon in *E. coli* (Tyler et al., 1967). Eine PFL

$$\frac{dx(t)}{dt} = \alpha \frac{x(t)^h}{k^h + x(t)^h}. \quad (9.2)$$

kann ebenfalls mit Hilfe der Hill-Funktion beschrieben werden. Genau wie in Gleichung (9.1) ist in Gleichung (9.2)  $k$  die Konstante der Rückkopplung,  $x(t)$  die Zielvariable,  $\alpha$  die Produktionsrate. Der Term  $(x(t)^h)/(k^h + x(t)^h)$  beschreibt die Aktivatoren, welche an den Promotor gebunden haben. Je mehr  $x(t)$  vorhanden sind, desto größer ist auch die Anzahl an gebundenen Aktivatoren  $(x(t)^h)/(k^h + x(t)^h)$  und  $h$  der Hill-Koeffizient. Mit steigender Konzentration von  $x$  geht die Produktionsrate gegen die maximale Produktionsrate  $\alpha$ .

Ist die PFL stark genug ausgeprägt, kann es zu zwei stabilen Zuständen kommen, also einem bistabilem Verhalten (siehe Kapitel 9.3). PFLs spielen dabei häufig bei Zelldifferenzierungsvorgängen eine große Rolle (Dubnau and Losick, 2006).

## 9.3. Bistabilität

Ein bistabiles System hat zwei stabil und einen instabilen Zustand. Das System tendiert zu einen der beiden stabilen Zuständen, welche stabile Gleichgewichtszustände sind. Ein Wechsel der Zustände kommt es durch Veränderung der Bedingungen. Innerhalb von biologischen System kann ein Wechsel der Zustände auch durch stochastische Fluktuationen (Ozbudak et al., 2004) zustandekommen.

### 9.3.1. Biologisches bistabiles Verhalten

Bistabiles Verhalten kommt in vielen Organismen vor, sowohl in Prokaryoten, als auch in Eukaryoten. In Prokaryoten konnte ein bistabiles Verhalten während der Sporulation von *Bacillus subtilis* gezeigt werden (Veening et al., 2005). Ebenfalls kommt es bei *B. subtilis* in der Regulation der genetischen Kompetenz zur Bistabilität (Maamar and Dubnau, 2005). Bei den Verhalten prägen ein Zellpopulationsheterogenität. Bei Eukaryoten wurde zum Beispiel ein bistabiles Verhalten bei der Oozyten-Reifung in *Xenopus laevis*, dem Afrikanischen Krallenfrosch, beobachtet (Xiong and Ferrell, 2003).

Alle bisher erläuterten bistabilen biologischen Systeme beinhalten eine positiven FBL (abgekürzt PFL). Diese PFL sind unterschiedlich in andere Systeme eingebettet und weisen unterschiedliche Komplexitäten und Strukturen auf. Die PFL ist häufig ein Zwei-Regulator-System (siehe Kapitel 2.6.1), kann jedoch auch eine einfache Autoregulation sein, bei der der Regulator seine eigene Produktion gesteuert wird. Diese Autoregulation findet man in der RovA-Regulation in *Y. pseudotuberculosis*. Die Bistabilität in bakteriellen Systemen bietet eine Art Schalter.

Innerhalb von *Y. pseudotuberculosis* zeigt RovA ein von der Temperatur abhängiges, bistabiles Verhalten (Herbst et al., 2009). Temperaturabhängige Bistabilität konnte bis jetzt beim  $\lambda$ -Phagen gezeigt werden. Die Temperaturregulation wird hier jedoch nicht über den Abbau, sondern nur über die Aktivität eines Regulators gesteuert (Bednarz et al., 2014).

Die RovA-Konzentration in *Y. pseudotuberculosis* wechselt zwischen fast null ( $>\approx 0$  nM) bei Wirts-Temperatur und ca. 35 nM bei niedriger Temperatur. Die Zeitskalen zwischen RovA-Aufbau und RovA-Abbau sind unterschiedlich, wodurch die Bistabilität entstehen kann. Der Aufbau von RovA dauert wesentlich länger als der Abbau.

Innerhalb dieser Arbeit wird die Bistabilität numerisch analysiert. Dies geschieht durch Berechnung der Gleichgewichtszustände für unterschiedliche Anfangsbedingungen. Diese Gleichgewichtszustände werden je nach Voraussetzung (z. B. Temperatur) miteinander verglichen.

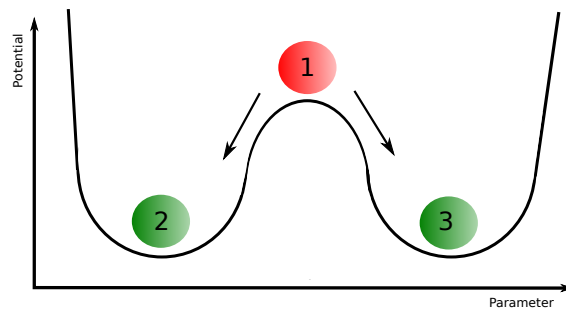
### 9.3.2. Potential

Das Potential  $U(r)$  beschreibt die potentielle Energie eines Systems. Mittels des Potentials kann das bistabile Verhalten von Systemen gut dargestellt werden. Ein bistabiles System hat in seinem Potential zwei lokale Minima, welche die stabilen Zustände sind und ein lokales Maximum, welches der instabile Zustand ist. Ist das Potential nicht minimal, so kann Energie abgegeben werden.

**Definition 11.** Ist die Funktion  $f(r)$  die rechte Seite einer Differentialgleichung nach  $r$ , so wird das Potential  $U(r)$  definiert durch

$$\frac{dU(r)}{dr} = -f(r) \tag{9.3}$$

In Abbildung 9.4 wird mittels des Potentials die Bistabilität verdeutlicht. Beide stabilen Zustände befinden sich bei den Minima (Kugel 2 und 3). Der instabile Zustand wird durch das Potential als Maximum dargestellt (Kugel 1). Das System kann dort verweilen, bei kleinsten Änderungen tendiert aber das gesamte System zu einem der stabilen Zustände.



**Abbildung 9.4.: Potential eines bistabilen Verhalten eines Systems**

Das Potential zeigt die potentielle Energie des Systems. Bistabiles Verhalten hat zwei stabile Zustände. Die lokalen Minima, Kugel 2 und 3, stellen die stabilen Zustände dar, das lokale Maximum (Kugel 1) einem instabilen Zustand. In diesem System tendiert die Variable in Richtung einer der beiden Minima, also der stabilen Zustände. Befindet sich das System in dem instabilen Zustand, so kann es auch dort verweilen. Gibt es jedoch geringste Veränderungen, so tendiert das System wieder in einen der stabilen Zustände.

## 9.4. Hysterese

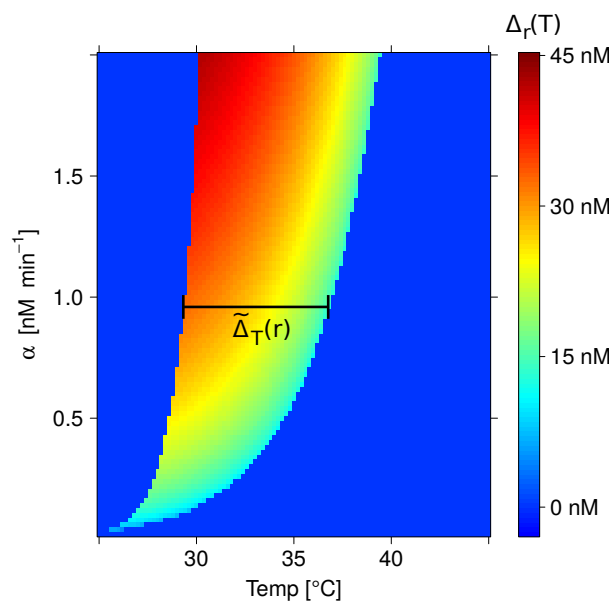
Ein System, das in mehr als nur einem Zustand vorkommen kann und dieser nicht nur von der Gegenwart sondern auch von seinen vorherigen Zuständen abhängt, besitzt eine Hysterese. Das Systemverhalten beruht nicht nur auf der derzeitigen Umwelt, sondern auch auf die vorherige.

**Definition 12. Hysterese** ist ein Systemverhalten, bei dem der Zustand des Systems nicht allein von den Eingangsgrößen abhängt, sondern auch von seinem vorherigen Zustand.

Innerhalb der Biologie kommt es zum Beispiel bei Zellwachstum zu Hysterese. So ist die Konzentration von Cyclin, welches den Zellzyklus beeinflusst, abhängig von vorherigen Bedingungen (Pomerening et al., 2003).

## 9.5. Phasendiagramme

Ein Phasendiagramm oder auch Zustandsdiagramm ist ein  $n$ -dimensionales (meist 2- oder 3-dimensionales) Diagramm, welches je nach Variablen unterschiedlichen Phasen, bzw. Zustände, darstellt. Die Zustände können zum Beispiel die Parameter sein, bei denen ein System bistabil ist (siehe Abbildung 9.5).



**Abbildung 9.5.: Phasendiagramm der Bistabilität des RovA-Modells zu der Temperatur und der maximalen Expressionsrate  $\alpha$**

Zwei stabile Gleichgewichtszustände werden für die RovA-Konzentration berechnet. Der Unterschied der beiden stabilen Zustände wird in diesem Phasendiagramm abgebildet. Innerhalb des dunkelblauen Bereiches ist die Phase bei dem ein stabiler Zustand der RovA-Konzentration vorhanden ist. Innerhalb des anderen Bereiches gibt es zwei verschiedene Zustände. In diesem Bereich wird durch die Farbskalierung der Unterschied der beiden Zustände gezeigt.

Genutzt werden in dieser Arbeit Phasendiagramme, um bei Veränderungen der Temperatur und der berechneten Parameter die Bistabilität anzuzeigen. Es wird nicht nur zwischen Bistabilität und Monostabilität unterschieden,

---

sondern es wird auch der Unterschied zwischen den beiden stabilen Zuständen dargestellt.

# **10. Modellierung der temperaturabhängigen RovA-Bistabilität**

Die RovA-Produktion zeichnet sich durch temperaturabhängige Bistabilität aus. Die Produktion wird mit einer positiven (PFL) und einer negativen (NFL) Feedback-Schleife gesteuert. Um zu zeigen welche Einflüsse die Temperatur, die einzelnen Parameter und die NFL haben, wurde ein deterministisches Modell erstellt. Die Parameter innerhalb dieses Modells wurden durch verschiedene experimentelle Daten gefittet.

Die RovA-Konzentrationsregulation soll mit einer gewöhnlichen Differentialgleichung (siehe Kapitel 2.7) abhängig von der Temperatur  $T$  dargestellt werden. Das Modell mit eingesetzten berechneten Parametern soll eine numerische Auswertung der Bistabilität und die Darstellung dieser mittels Phasendiagrammen erlauben.



## 10.1. RovA-Produktion als Differentialgleichung

Die Konzentration von RovA ( $r(t)$ ) wird mittels einer Differentialgleichung (Gleichung (10.1)) dargestellt (siehe Kapitel 2.7). Die Differentialgleichung

$$\dot{r}(t) = f(r, T, p, t) \quad (10.1)$$

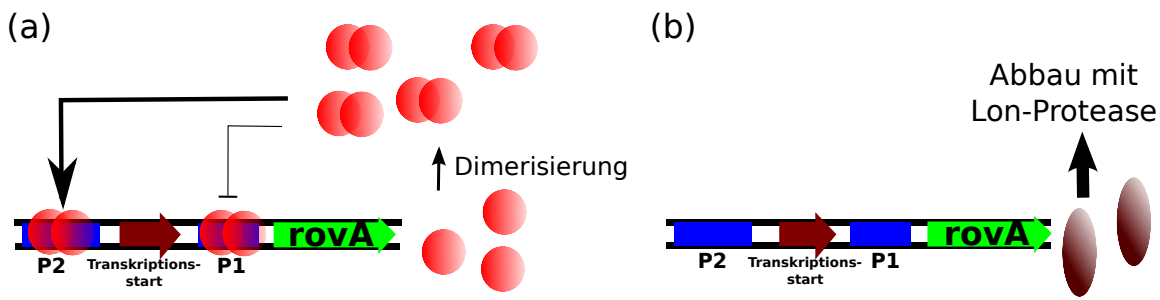
hängt von der RovA-Konzentration selbst ( $r(t)$ ), der Zeit  $t$ , der Temperatur ( $T$ ) und von verschiedenen durch Experimente bestimmbar Parametern  $p$  ab.

Die zeitliche Veränderung der RovA-Konzentration  $\dot{r}$  ist abhängig von den Parametern  $p$ . In Abbildung 10.1 wird das Operon von RovA abgebildet und dargestellt, welche Einflüsse auf das RovA-System möglich sind.

RovA wird auf- und wieder abgebaut. Der Aufbau wird durch  $A_{\text{RovA}}$  und die Abbaurate durch  $\delta$  dargestellt. Die Produktion von RovA

$$\dot{r}(t) = A_{\text{RovA}} - \delta(T) \cdot r. \quad (10.2)$$

wird in die positive Produktion ( $A_{\text{RovA}}$ ) und den negativen Abbau  $\delta$  unterteilt. Bei 37°C ist im Gegensatz zu 25°C der Abbau von RovA stärker ausgeprägt. Der Abbau ist temperaturabhängig und wird mit  $\delta(T)$  beschrieben. Je mehr RovA  $r$  vorhanden ist, um so mehr kann abgebaut werden  $\delta(T) \cdot r$ .



**Abbildung 10.1.: RovA-Operon bei 25°C (a) und 37°C (b):**

Der Promotor für die RovA-Produktion wird für 25°C, wie in Abbildung 9.1, und für 37°C, wie in Abbildung 9.2, gezeigt. RovA (rote Kugeln/ dunkelrote Ellipsen) bindet bei 25°C ist die an den Promotoren  $P_1$  und  $P_2$  stärker als bei 37°C. Jedoch findet der Abbau von RovA bei 37°C wesentlich stärker statt.

Die Produktion von RovA ( $A_{\text{RovA}}$ ) besitzt eine maximale Produktionsrate  $\alpha$  und eine basale Expressionsrate von  $\alpha_0$  (Yagil and Yagil, 1971). An dem aktivierenden Promotor bindet RovA mit einer starken Bindungskonstante von  $k_a(T)$ , an den reprimierenden Promotor mit einer schwachen Bindungskonstante von  $k_r(T)$ , wobei  $k_a(T)$  und  $k_r(T)$  die Konstanten der Rückkopplung der beiden FBLs darstellen (siehe Kapitel 9.2). Sowohl  $k_a(T)$  als auch  $k_r(T)$  hängen direkt von der Temperatur  $T$  ab. In Abbildung 10.1 (a) wird gezeigt, dass bei 25°C RovA mit seiner Konformation als Dimer vorliegt und an die Promotoren mit  $k_a(T)$  und  $k_r(T)$  bindet. Bei 37°C ändert sich die Konformation von RovA (Abbildung 10.1 (b)), wodurch es keine Dimere bildet und die Bindekonstante sehr viel geringer sind als bei 25°C.

Mit der Gleichungen (9.1) und (9.2) (Kapitel 9.2), in denen die beiden FBLs beschrieben werden, und mit der Hill-Funktion (Kapitel 2.7) wird die Differentialgleichung dargestellt. Die Bindestellen für die NFL und für die PFL liegen innerhalb eines Operon (siehe Abbildung 2.1). Ist die NFL vollständig aktiviert oder ist die PFL vollständig deaktiviert, wird kein RovA produziert. Daher beeinflussen beide Schleifen gleich stark als Produkt das Wachstum

von RovA. Die Bindungsgleichungen für die Schleifen werden miteinander multipliziert. Ebenfalls mit in das Modell gehen die maximale Produktionsrate  $\alpha$  und die basale Expressionsrate  $\alpha_0$  ein. Der Hill-Koeffizient wird hier  $h_a$  für den aktivierenden und  $h_r$  für die reprimierende FBL genannt. Die RovA-Produktion

$$\dot{r}(t) = \alpha_0 + \alpha \cdot \frac{r^{h_a}}{k_a(T)^{h_a} + r^{h_a}} \cdot \frac{k_r(T)^{h_r}}{k_r(T)^{h_r} + r^{h_r}} - \delta(T) \cdot r. \quad (10.3)$$

beinhaltet Parameter und Variablen, die in Tabelle 10.1 dargestellt sind.

**Tabelle 10.1.: Involvierte Parameter in die Modellierung der RovA Produktion:**

Auflistung aller Parameter und Variablen, welche in das RovA-System mit eingehen.

Parameter	Funktion
$r(t)$	RovA-Konzentration (in nM)
$t$	Zeit (in min)
$T$	Temperatur (in °C)
$\alpha_0$	Basale Expressionsrate der RovA-Konzentration (in nM/min)
$\alpha$	Max. Expressionsrate der RovA-Konzentration (in nM/min)
$k_a(T)$	Bindekonstante RovAs an akti. Promotor $P_2$ , abhängig von $T$
$k_r(T)$	Bindekonstante RovAs an repr. Promotor $P_1$ , abhängig von $T$
$h_a$	Hill-Koeffizient des aktivierenden Promotors $P_2$
$h_r$	Hill-Koeffizient des reprimierenden Promotors $P_1$
$\delta(T)$	Abbaurrate von RovA, abhängig von $T$ (in 1/min)

Die drei Experimentreihen werden genutzt, um die Parameter aus Tabelle 10.1 zu bestimmen.

## 10.2. Temperaturabhängige Abbaurate von RovA

Die temperaturabhängige Abbaurate  $\delta(T)$  von RovA wurde für fünf verschiedenen Temperaturen bestimmt (25, 28, 31, 34 und 37°C), (Herbst et al., 2009). *Y. pseudotuberculosis* wurde auf LB-Medium (engl. lysogeny broth, komplexes Nährmedium für Bakterienkultur) bei jeder der fünf Temperaturen bis zur exponentiellen Wachstumsphase gezogen. Die Proteinsynthese wurde innerhalb dieser Wachstumsphase durch Zugabe von Chloramphenicol gestoppt. Chloramphenicol ist ein Translationshemmer, der die Bindung der Ribosomen an die DNA hemmt. Nach 0 min, 30 min, 60 min und 90 min wurde je ein Western-Blot mit diesen Proben durchgeführt. Bei dem Western-Blot werden die mittels der Agarose-Gelelektrophorese aufgetrennten Proteine auf eine Trägermembran übertragen. Die durch den Western-Blot analysierten Banden von Proteinen wurden mittels des Auswerteprogramms ImageJ<sup>1</sup> verarbeitet und miteinander verglichen (Schneider et al., 2012).

Über eine nichtlineare Regression konnten die Abbauraten für die fünf Temperaturen bestimmt werden. Der Abbau von RovA ist exponentiell. Für die nichtlineare Regression wurde die Abbaugleichung  $r = r_0 \cdot e^{(-\delta \cdot t)}$  genutzt. Die Experimente ergaben die Daten in Tabelle 10.2. Die Daten zeigen, dass die Abbaurate mit größer werdender Temperatur ebenfalls ansteigen. Der exponentiell ansteigende Abbau von RovA wird durch die Lon-Protease gesteuert.

---

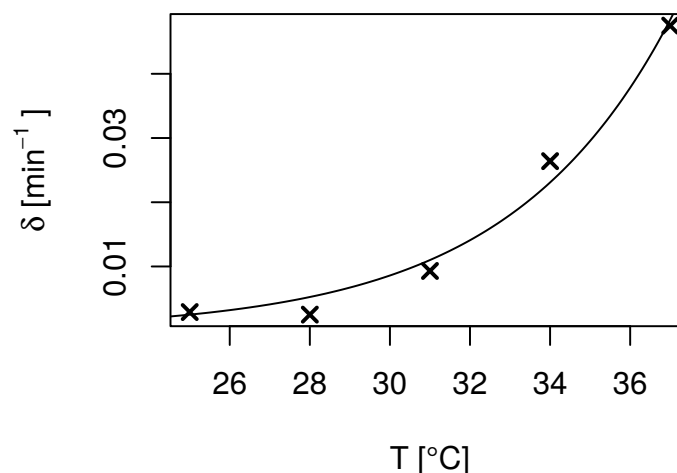
<sup>1</sup><http://imagej.nih.gov/ij/>

**Tabelle 10.2.: Abbauraten von RovA für verschiedene Temperaturen**

Mittels einer nichtlinearen Regression werden die Abbaurate von RovA für fünf verschiedene Temperaturen zu der Funktion  $r = r_0 \cdot e^{(-\delta \cdot t)}$  berechnet. Für 25°C, 28°C, 31°C, 34°C und 37°C werden die Abbauraten  $\delta(T)$  mittels einer nichtlinearen Regression beschrieben (Schneider et al., 2012).

Temperatur [°C]	$\delta(T)$ [min <sup>-1</sup> ]
25	0.0029
28	0.0025
31	0.0093
34	0.0264
37	0.0475

Aufbauend auf den Daten aus Tabelle 10.2 wurden eine lineare und eine nichtlineare Regressionen mittels Minimierung der Abweichungsquadrate (Residual Sum of Square, RSS) durchgeführt, um den Zusammenhang von Abbaurate und Temperatur zu ermitteln. Die Abweichung der Messungen zu der Regressionskurven wurden quadriert, aufsummiert und dann minimiert.



**Abbildung 10.2.: Abbauraten von RovA in Abhängigkeit von der Temperatur:** Für 25°C, 28°C, 31°C, 34°C und 37°C werden die Abbauraten  $\delta(T)$  aufgetragen. Der Abbau steigt mit höher werdender Temperatur. Die eingezeichnete Regressionskurve ist die mit den Daten durchgeführte nicht-lineare Regression aus Gleichung (10.4).

Zu beiden Regressionen wurden die RSS berechnet. Zu den Daten aus Tabelle 10.2 wurde ein RSS von 0.0002052 mit der linearen Regression und ein RSS von 0.00002245 mit der nichtlinearen Regression mit der Funktion  $\delta(T) = \delta_0 \cdot e^{(\delta_1 \cdot T)}$  berechnet.

Durch den niedrigeren RSS wird die nichtlineare Funktion genutzt. Die Funktion

$$\delta(T) = 5.2 \cdot 10^{-6} \cdot e^{(0.2469 \cdot T)} \quad (10.4)$$

wird in Abbildung 10.2 dargestellt.

### 10.3. Temperaturabhängige DNA-Bindungskonstanten von RovA

Die Bindungen von RovA an die Bindestellen der beiden Promotoren, den aktivierenden und den reprimierenden, sind ebenfalls temperaturabhängig (siehe Kapitel 9.1.2). Die Bindekonstante für den aktivierenden Promotor wird mit  $k_a(T)$  und für den reprimierenden mit  $k_r(T)$  bezeichnet. Mittels einem Electrophoretic Mobility Shift Assay (EMSA) wird die Bindung zwischen der DNA der Promotoren und dem RovA-Protein gemessen. EMSA, auch Band Shift Assay genannt, ist eine Affinitätselektrophorese von RNA- oder DNA-bindenden Proteinen.

Ein EMSA wurde je für beide Promotoren, dem aktivierenden und dem reprimierenden, durchgeführt bei zwei unterschiedlichen Temperaturen (Klein, J., unpubliziert). Hierbei wird eine immer größer werdende RovA-Konzentration

titriert und die Konzentration des Protein-DNA-Komplexes gemessen. Daraus resultieren zwei sigmoide Bindungskurven, jeweils für den aktivierenden und den reprimierenden Promotor.

Da die Bindung temperaturabhängig ist, wird die Bindekonstante für zwei unterschiedliche Temperaturen berechnet. Es wurden dem entsprechend für 25°C und 37°C für beide Promotoren ein EMSA durchgeführt (siehe Abbildung 10.3). Es wurden insgesamt vier Electrophoretic Mobility Shift Assays (EMSA) durchgeführt.

Die Bindekonstanten  $k_a(T)$  und  $k_r(T)$  werden mittels nichtlinearen Regressionen mit der Funktion

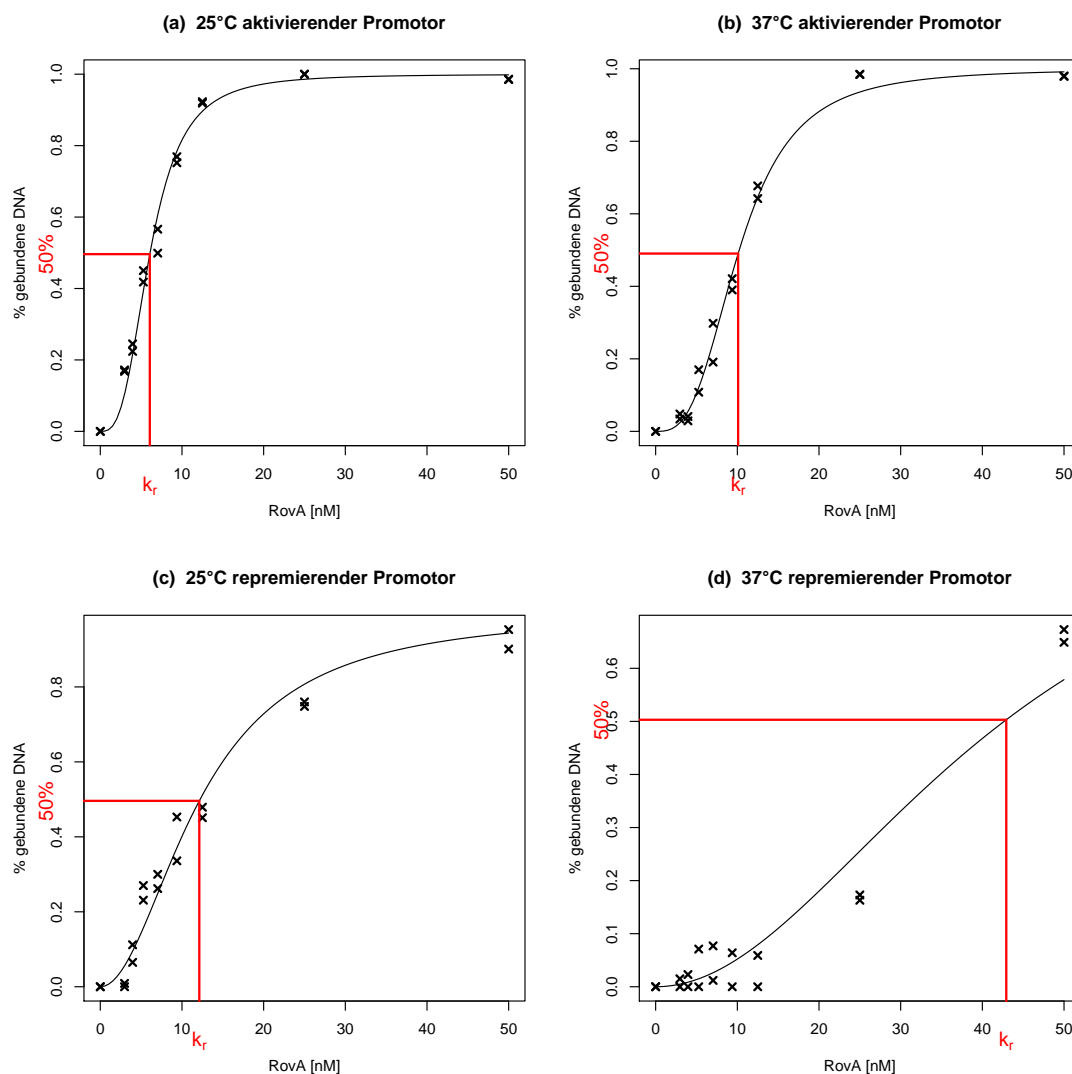
$$\nu_a = \frac{r^{h_a}}{r^{h_a} + k_a^{h_a}}, \text{ bzw. } \nu_r = \frac{r^{h_r}}{r^{h_r} + k_r^{h_r}}$$

passend zu den Daten der EMSA-Experimente durchgeführt. Dies geschieht über die Minimierung der Abweichungsquadrate (Residual Sum of Sqaure, RSS). Die prozentuale Anzahl der gebundenen DNA wird mit  $\nu_a$ , bzw.  $\nu_r$  bezeichnet,  $r(t)$  ist die Konzentration des Proteins RovA und  $h_a$ , bzw.  $h_r$  bezeichnen die Hill-Koeffizienten. Die Daten und die Regressionskurven sind in Abbildung 10.3 dargestellt.

Durch diese Regressionskurven können die Halbbindekonstante von RovA an die Promotoren berechnet werden. Diese gibt an, wie groß die Bindung zwischen DNA (also den Promotoren) und Protein ist. Ist 50 % der DNA gebunden, so spiegelt die RovA-Konzentration den Betrag der Bindekonstante wider. In Abbildung 10.3 wird dieses durch die roten Linien gekennzeichnet.

Die Regression hängt direkt von den Angaben des Hill-Koeffizienten  $h_a$ , bzw.  $h_r$  ab und wird zur näheren Untersuchung auf drei unterschiedliche Weisen beschrieben. Zum einen wird der Hill-Koeffizient variabel gelassen und dadurch von der Regression mit berechnet. Dadurch kann es zu

einem Hill-Koeffizient kommen, der nicht zu den natürlichen Zahlen gehört ( $h \notin \mathbb{N}$ ). Zur Vereinfachung wurden zwei weitere Regressionen für jeden Promotor bei jeder Temperatur mit den Hill-Koeffizienten von  $h = 2$  und  $h = 3$  durchgeführt (siehe Tabelle 10.3).



**Abbildung 10.3.: Electrophoretic Mobility Shift Assay für beide Promotoren der RovA-Produktion bei zwei Temperaturen (Klein, J., unpubliziert):**

Für den aktivierenden (a) und (b) und den reprimierenden (c) und (d) Promotor werden je ein EMSA bei 25°C (a) und (c) und 37°C (b) und (d) durchgeführt. In rot ist der Wert zu erkennen, bei dem 50 % der DNA gebunden wird und somit die Bindekonstanten  $k_a(T)$  und  $k_r(T)$  widerspiegelt werden.



Um eine temperaturabhängige Funktion von der reprimierenden und aktivierende Bindekonstante zu erstellen, ist es notwendig, dass der Hill-Koeffizient für beide Temperaturen gleich ist ( $h_a(25^\circ \text{ C}) = h_a(37^\circ \text{ C})$  und  $h_r(25^\circ \text{ C}) = h_r(37^\circ \text{ C})$ ), damit das Verhalten der Kooperativität bei unterschiedlicher Temperatur pro Promotor gleich ist. Es ist nicht notwendig, dass der Hill-Koeffizient für den aktivierenden und den reprimierenden Promotor gleich ist, da es sich um zwei unterschiedliche Promotoren handelt und somit auch die Kooperativität und damit der Hill-Koeffizient unterschiedlich stark sein kann. Die Ergebnisse werden in Tabelle 10.3 gezeigt.

**Tabelle 10.3.:** Berechnung der Bindekonstanten für den aktivierenden und reprimierenden RovA-Promotor bei  $25^\circ \text{ C}$  und  $37^\circ \text{ C}$ , je mit variablen Hill-Koeffizient oder Hill-Koeffizient  $h = 2$  oder  $h = 3$ . Es werden immer die fünf Datenpunkte aus Tabelle 10.2 genutzt, daher ist bei allen RSS-Werten  $n = 5$ .

Promotor	Temp.	Bindungskonst. nM	Hillkoef.	RSS
aktivierend	$25^\circ \text{ C}$	$k_a(T)$ 6.07 nM	$h_a$ 2.62	0.02203
aktivierend	$25^\circ \text{ C}$	$k_a(T)$ 6.01 nM	$h_a$ 2	0.05274
aktivierend	$25^\circ \text{ C}$	$k_a(T)$ 6.09 nM	$h_a$ 3	0.02995
aktivierend	$37^\circ \text{ C}$	$k_a(T)$ 10.19 nM	$h_a$ 3.17	0.01713
aktivierend	$37^\circ \text{ C}$	$k_a(T)$ 10.69 nM	$h_a$ 2	0.09329
aktivierend	$37^\circ \text{ C}$	$k_a(T)$ 10.23 nM	$h_a$ 3	0.01808
reprimierend	$25^\circ \text{ C}$	$k_r(T)$ 12.61 nM	$h_r$ 1.72	0.03646
reprimierend	$25^\circ \text{ C}$	$k_r(T)$ 12.22 nM	$h_r$ 2	0.04782
reprimierend	$25^\circ \text{ C}$	$k_r(T)$ 11.45 nM	$h_r$ 3	0.16690
reprimierend	$37^\circ \text{ C}$	$k_r(T)$ 40.66 nM	$h_r$ 3.15	0.01633
reprimierend	$37^\circ \text{ C}$	$k_r(T)$ 42.67 nM	$h_r$ 2	0.04487
reprimierend	$37^\circ \text{ C}$	$k_r(T)$ 40.69 nM	$h_r$ 3	0.01665

Man kann an Abbildung 10.3 (d) erkennen, dass das Bindungsverhalten von RovA an den reprimierenden Promotor bei  $37^\circ \text{ C}$  nicht spezifisch ist, also die Messungen für niedrige RovA-Konzentrationen einer Punktwolke gleichen. Daher ist die Bindung zwischen RovA und dem reprimierendem Promotor bei dieser Temperatur so gering, dass die Bindekonstante sehr groß

sein muss. Die Daten für dieses Bindeverhalten sollten für die Wahl des Hill-Koeffizienten einen geringeren Einfluss haben.

Genutzt werden soll der Koeffizient mit dem niedrigsten RSS (Tabelle 10.3). Es wird der Hill-Koeffizient  $h_a = 3$  für den aktivierenden Promotor und  $h_r = 2$  für den reprimierenden Promotor genutzt. Es sind je zwei Bindekonstanten

$$\begin{aligned} k_a(25^\circ\text{C}) &= 6.09 \text{ nM} \\ k_a(37^\circ\text{C}) &= 10.23 \text{ nM} \\ k_r(25^\circ\text{C}) &= 12.22 \text{ nM} \\ k_r(37^\circ\text{C}) &= 42.67 \text{ nM} \end{aligned} \tag{10.5}$$

pro Promotor gegeben.

Die Abhängigkeit der Bindekonstanten von der Temperatur kann linear ( $k = \lambda + \mu \cdot T$ ) oder nichtlinear (hier exponentiell  $k = \lambda \cdot e^{(\mu \cdot T)}$ ) sein, siehe Abbildung 10.4.

Ein linearer Zusammenhang von Bindekonstante und Temperatur ist

$$\begin{aligned} k_a &= -2.535 + 0.345 \cdot T, \\ k_r &= -51.2175 + 2.5375 \cdot T. \end{aligned}$$

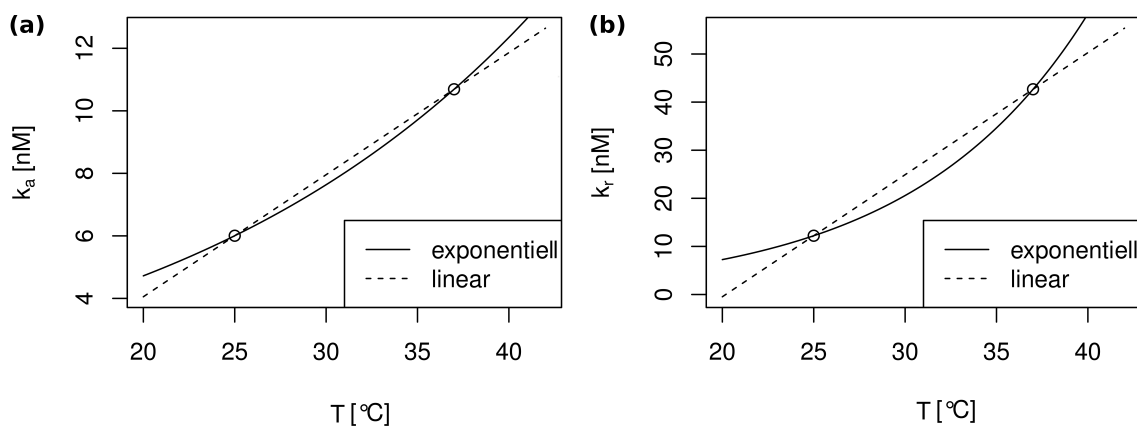
Mit der nichtlinearen Anpassung werden die Parameter von  $k_a(T) = \lambda_a \cdot e^{(\mu_a \cdot T)}$  und  $k_r(T) = \lambda_r \cdot e^{(\mu_r \cdot T)}$  berechnet. Genutzt werden dazu die Werte aus Gleichung (10.5). Somit kann man die Parameter  $\mu_a = 0.0432/^\circ\text{C}$ ,  $\lambda_a = 2.0669 \text{ nM}$ ,  $\mu_r = 0.1042/^\circ\text{C}$  und  $\lambda_r = 0.9031 \text{ nM}$  berechnen und in die Bin-

dungsgleichungen der Promotoren

$$k_a(T) = 2.0669 \cdot e^{(0.0432 \cdot T)} \quad (10.6)$$

$$k_r(T) = 0.9031 \cdot e^{(0.1042 \cdot T)} \quad (10.7)$$

eingehen.



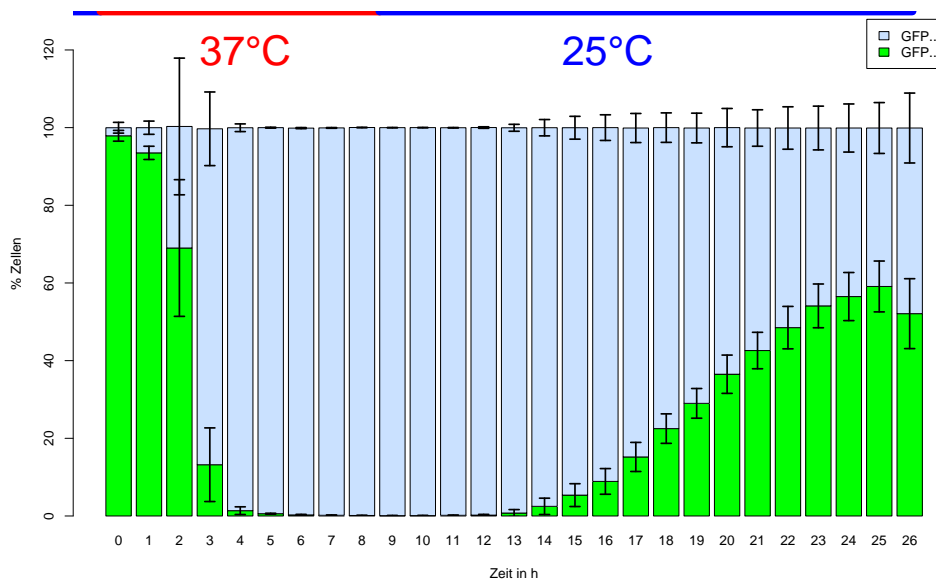
**Abbildung 10.4.: Temperaturabhängigkeit der Bindekonstanten**

Temperaturabhängigkeit der aktivierenden Bindekonstanten  $k_a(T)$  (a) und der reprimierenden Bindekonstante  $k_r(T)$  (b) mit linearer und nichtlinearer Regression.

Der lineare Zusammenhang der Bindekonstante von der Temperatur erscheint bei nur zwei gemessenen Punkten zunächst sinnvoll. Jedoch ist dieser Zusammenhang eingeschränkt auf einen gewissen Temperaturbereich. Bei Temperaturen bei  $T < 7.347826^{\circ}\text{C}$  kommt es zu einer negativen Bindekonstante  $k_a$  (siehe Abbildung 10.4). Betrachtet man die Bindekonstante  $k_r(T)$ , so wird diese bereits bei Temperaturen von  $T < 20.18424^{\circ}\text{C}$  negativ. Eine Bindekonstante an den Promotoren kann nicht negativ werden, daher wird für die Bindekonstante ein nichtlinearer Zusammenhang angenommen. Ebenfalls kommt es bei immer höheren Temperaturen zu exponentiell wachsenden Bindekonstanten, die Bindung wird also immer unspezifischer. Dieses spiegelt das Systemverhalten besser wider als der lineare Zusammenhang.

## 10.4. Temperaturwechsel Experiment

Das dritte Experiment für die Berechnung der Parameter des RovA-Modells ist ein Temperaturwechsel-Experiment (Klein, J., Nuss, A., unpubliziert). Die RovA-Promotor-Aktivität wurde mittels eines RovA-Promotor-Reporter-System gemessen.



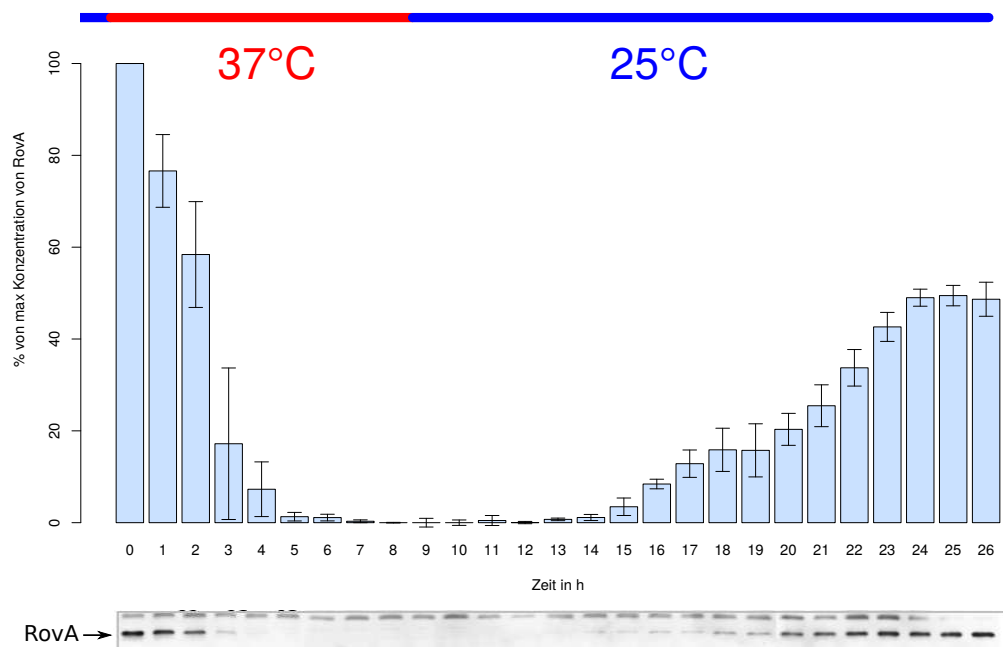
**Abbildung 10.5.: Temperaturwechsel Experiment mit einem RovA-Promotor-Reporter-System (Klein, J., Nuss, A., unpubliziert)**

Auf der x-Achse aufgetragen wird die Zeit in Stunden  $h$  und die jeweilige Temperatur markiert. Auf der y-Achse ist der prozentuale Anteil an Zellen, welche GFP enthielten, mit grün dargestellt. Mit hellblau der Anteil der Zellen ohne GFP. Bei niedrigen Temperaturen in der Vorkultur sind nahezu 100 % der Zellen grün, zeigen also eine RovA-Promotor-Aktivität. Die Vorkultur wird bei 25°C gezogen. Bei 37°C geht die Promotoraktivität zurück und steigt wieder bei erneuten herabsetzen der Temperatur auf 25°C an.

Es wurden beide RovA-Promotoren mit einem Reporter-Gen auf einem Plasmid mit geringer Kopienanzahl fusioniert. Als Reporter-Gen wurde das Gen

für GFP genutzt. GFP ist ein grün fluoreszierendes Protein (engl. green fluorescent protein), dass bei Anregung mit blauem oder ultraviolettem Licht grün fluoresziert (Tsien, 1998).

Das Plasmid wurde in den Wildtyp von *Y. pseudotuberculosis* transferiert. Die Aktivität des RovA-Promotors konnte direkt an der GFP-Konzentration und an der grünen Lichtintensität betrachtet werden. Der Wildtyp mit dem Plasmid wurde in einer Vorkultur bei 25°C angezogen.



**Abbildung 10.6.: Temperaturshift Experiment mit Western-Blot (Klein, J., Nuss, A., unpubliziert)**

Auf der x-Achse wird die Zeit in Stunden (*h*) gezeigt und die Temperatur (25°C oder 37°C). Auf der y-Achse wird die Konzentration von RovA im Bezug auf die höchste Konzentration im Western-Blot dargestellt (höchste Konzentration ist somit bei 100 %). Im unterem Teil der Abbildung ist ein Beispiel des Western-Blots aufgetragen. Man kann hier wie in Abbildung 10.5 erkennen, dass die Konzentration von RovA bei hohen Temperaturen abfällt und nur langsam bei niedrigeren Temperaturen wieder aufgebaut wird.

Über einen Zeitraum von acht Stunden bei 37°C wurde die Intensität von GFP gemessen. Danach folgte ein Temperaturwechsel zu 25°C und es wurde erneut die GFP-Intensität über 18 Stunden gemessen. Die GFP-Stärke wurde stündlich ermittelt (Abbildung 10.5).

Um diese Daten zu verifizieren, wurde in dem Temperaturwechsel-Experiment die RovA-Konzentration ebenfalls pro Stunde direkt mittels eines Western-Blots bestimmt (Klein, J., Nuss, A., unpubliziert). Hier wurde nicht die RovA-Promotor-Aktivität bestimmt, sondern direkt die Konzentration von RovA. In Abbildung 10.6 wird das Ergebnis gezeigt. Es ist deutlich zu erkennen, dass dieses Experiment des RovA-Verhalten exakt dem des Promotors von Abbildung 10.5 widerspiegelt. Beide Experimente wurden dreimal wiederholt und spiegelte alle das in Abbildung 10.5 und Abbildung 10.6 dargestellte Verhalten wider.

### 10.4.1. Statistische Berechnung der Expressionsraten

Die Parameter der Promotorbindung und der Abbaurate wurden bereits berechnet. Damit das Modell exakt das RovA-Verhalten widerspiegelt, müssen die basale Expressionsrate  $\alpha_0$  und die maximale Expressionsrate  $\alpha$  der RovA-Produktion bestimmt werden. Mittels des Temperaturwechsel-Experiments aus Kapitel 10.4 können die Expressionsraten statistisch bestimmt werden.

Es konnten beide Raten berechnet werden (Müller, J., unpubliziert). Dazu wurde zwei-dimensional für die variierende Raten ( $\alpha_0$  und  $\alpha$ ) die Abweichungsquadrate (RSS - aus Kapitel 10.3) berechnet. Bei minimalen Abwei-

chungsquadraten ergeben sich folgende Raten:

$$\alpha = 0.7110374 \text{ nM/min} \quad (10.8)$$

$$\alpha_0 = 0.0018724 \text{ nM/min} \quad (10.9)$$

### Anzahl an RovA Molekülen pro Zelle

Bei 25°C konnten ca. 310 RovA Moleküle pro Zelle ( $m = 310$ ) nachgewiesen werden (Nuss, A., unpubliziert). Geht man davon aus, dass eine *Y. pseudotuberculosis*-Zelle eine Länge von circa  $l_y = 3\mu\text{m}$  und eine Breite von  $w_y = 1\mu\text{m}$  besitzt und das Bakterium eine Stäbchen-Form hat, so wäre das Volumen  $V_y = (0.5w_y)^2 \cdot \pi \cdot l = 2.35619\mu\text{m}^3 = 2.35619 \cdot 10^{-15} \text{ Liter}$  (siehe Kapitel 9.1). Die Konzentration von RovA in der Zelle bei 25°C ist somit

$$C = \frac{m}{V_y} = 1.315683 \cdot 10^{17} \frac{\text{Moleküle}}{\text{Liter}}$$

oder in Molarität angegeben

$$\text{Mol} = \frac{C}{6.022 \cdot 10^{23}} = 2.18479 \cdot 10^{-7} \text{ M} = 218.479 \text{ nM}$$

Durch die Parameter  $\alpha = 0.7110374 \text{ nM/min}$  und  $\alpha_0 = 0.0018724 \text{ nM/min}$  erwartet man eine Konzentration von 35nM RovA pro Zelle. Dieses kann leicht durch den Unterschied zwischen freiem und gebundenem RovA erläutert werden. Ebenso kommt RovA als Monomer und als Dimer vor. Innerhalb der Messung der Anzahl an Molekülen wurden sowohl Dimere als auch Monomere gemessen, im RovA-Regulationsmodell kann RovA nur als Dimer agieren.

## 10.5. Gesamt-Modell

Durch diese Experimente kann ein Modell mit allen Parametern aufgestellt werden. Wie in schon Gleichung (10.3) hat das Modell folgenden Grundaufbau:

$$\dot{r}(t) = \alpha_0 + \alpha \cdot \frac{r^{h_a}}{k_a(T)^{h_a} + r^{h_a}} \frac{k_r(T)^{h_r}}{k_r(T)^{h_r} + r^{h_r}} - \delta(T) \cdot r. \quad (10.10)$$

Die Parameter sind in Tabelle 10.4 beschrieben und konnten durch die Daten aus den drei Experimentreihen berechnet werden.

**Tabelle 10.4.: Alle Parameter des RovA-Modells**

Parameter	Funktion	Werte
$r(t)$	RovA-Konzentration (in nM)	
$t$	Zeit (in min)	
$T$	Temperatur (in °C)	
$\alpha_0$	Basale Expressionsrate	0.0018724 nM/min
$\alpha$	Maximale Expressionsrate	0.7110374 nM/min
$k_a(T)$	Bindekonstante von RovA an $P_2$	$2.0669 \cdot e^{(0.0432 \cdot T)}$
$k_r(T)$	Bindekonstante von RovA an $P_1$	$0.9031 \cdot e^{(0.1042 \cdot T)}$
$h_a$	Hill-Koeffizient des akt. Promotors	3
$h_r$	Hill-Koeffizient des repr. Promotors	2
$\delta(T)$	Abbaurrate von RovA	$5.2 \cdot 10^{-6} \cdot e^{(0.2469 \cdot T)}$



# 11. Ergebnisse und Diskussion

Die RovA-Produktion wurde mittels einer Differentialgleichung dargestellt. Die Parameter konnten in Kapitel 10 bestimmt werden. Die Einflüsse der einzelnen Parameter und der Autoregulationen der RovA-Produktion werden in Eigenarbeit untersucht.

## 11.1. Charakterisierung der RovA-Promotoren

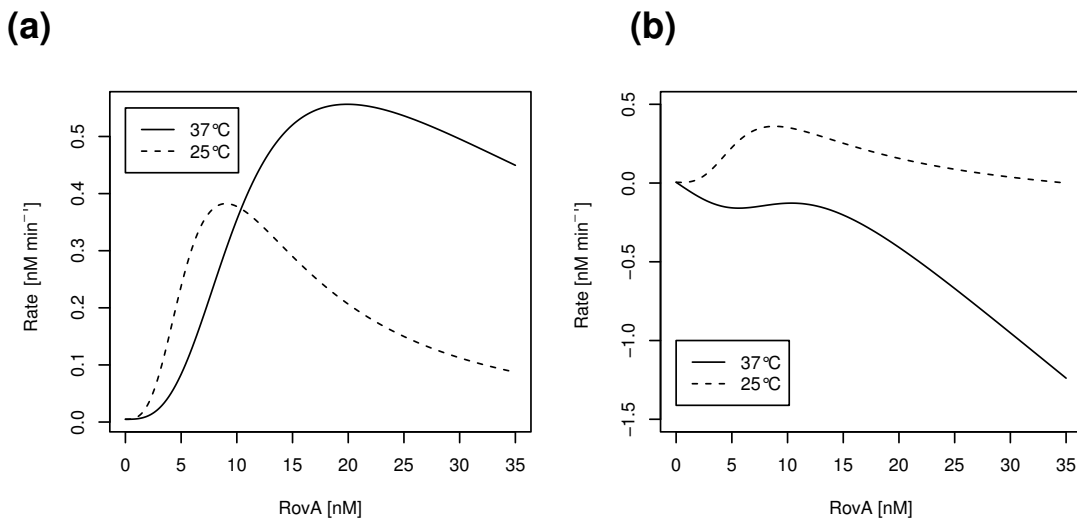
Mit dem Modell aus Gleichung (10.10) und den dazugehörigen Parametern aus Tabelle 10.4 kann die Produktion von RovA analysiert werden.

Abhängig von der Temperatur  $T$  und der RovA-Konzentration ( $r(t)$ ) ist der Promotor aktiv. Somit findet auch abhängig von der Temperatur und der RovA-Konzentration die Produktion von RovA statt. Die Produktion von RovA wird mittels der rechten Seite der Differentialgleichung (10.10) berechnet. Die gekoppelte Aktivität der beiden RovA-Promotoren  $A_{\text{RovA}}$  wird definiert als Produktion von RovA ohne Abbau, daher ist diese

$$A_{\text{RovA}} = \alpha_0 + \alpha \cdot \frac{r^{h_a}}{k_a(T)^{h_a} + r^{h_a}} \frac{k_r(T)^{h_r}}{k_r(T)^{h_r} + r^{h_r}}$$

(siehe Gleichung (10.2)).

In Abbildung 11.1 (a) wird abhängig von der RovA-Konzentration die Aktivität des Promotors durch die Funktion  $A_{\text{RovA}}$  dargestellt. Die Promotoraktivität hat eine von der Temperatur abhängige Steigung. Bei niedrigen Temperaturen ( $25^\circ\text{C}$ ) steigt die Promotoraktivität schon bei kleinen RovA-Konzentrationen schnell an. Die Produktionsstärke von RovA sinkt jedoch bei einer RovA-Konzentration von ca. 7 nM. Bei höheren Temperaturen (ab  $37^\circ\text{C}$ ) liegt die maximale Promotoraktivität bei ca. 20 nM. Ab dieser Konzentration (20 nM bei  $37^\circ\text{C}$  und 7 nM bei  $25^\circ\text{C}$ ) wird der Einfluss der reprimierenden Bindestelle des Promotors größer als der Einfluss der aktivierenden. Der Einfluss der reprimierenden Bindestelle bei  $37^\circ\text{C}$  ist erst ab höheren RovA-Konzentrationen vorhanden und auch sehr viel geringer, da die Bindung an den reprimierenden Promotor bei  $37^\circ\text{C}$  fast inaktiv ist (siehe Kapitel 10.3).



**Abbildung 11.1.: Promotor-Aktivität und Produktionsfunktion von RovA.** (a) Promotor-Aktivität von RovA als Funktion  $A_{\text{RovA}}$  abhängig von RovA ohne Abbaurate bei  $25^\circ\text{C}$  und  $37^\circ\text{C}$ . (b) Gesamte Produktion  $A_{\text{RovA}} - \delta(T)$  von RovA mit temperaturabhängiger Abbaurate  $\delta(T)$  bei  $25^\circ\text{C}$  und  $37^\circ\text{C}$ . Trotz der größeren Promotoraktivität von RovA bei  $37^\circ\text{C}$  ist die gesamte Produktionsrate negativ, da hier die Abbaurate mit involviert ist.

In Abbildung 11.1 (a) wird gezeigt, dass eine höhere Temperatur auch eine höhere RovA-Promotoraktivität hervorrufen. Dieses resultiert jedoch nicht in eine höhere RovA-Konzentration. Die involvierte Abbaurate  $\delta(T)$  hat bei höheren Temperaturen ebenfalls einen größeren Einfluss und gleicht damit die Promotoraktivität aus.

In Abbildung 11.1 (b) ist die absolute Produktion von RovA mit  $A_{\text{RovA}} - \delta(T)$  aufgetragen. Diese ist eine Kombination aus der Promotoraktivität und den Abbau von RovA  $\delta(T)$ . Hier kann man erkennen, dass der Abbau von RovA die erhöhte Produktion bei höheren Temperaturen und höheren RovA-Konzentrationen wieder aufhebt. Die Gesamt-Produktion von RovA ist bei höheren Temperaturen immer negativ. Dieses erfolgt durch den sehr großen Einfluss der Lon-Protease, welche erst bei höheren Temperaturen aktiv wird (siehe Kapitel 9.1.2).

## 11.2. Analyse der Bistabilität

Der Gleichgewichtszustand der RovA-Konzentration lässt sich durch das gegebene Modell numerisch berechnen. Die RovA-Konzentration hat für jede Temperatur mindestens einen Gleichgewichtszustand. Innerhalb eines bestimmten Temperaturbereiches existieren zwei stabile Gleichgewichtszustände. Innerhalb dieses Temperaturbereich ist die RovA-Konzentration bistabil. Der Gleichgewichtszustand ist abhängig von der Temperatur. Bei höheren Temperaturen ist die RovA-Konzentration kleiner als bei niedrigen Temperaturen, erkennbar an der Produktionsgröße in Abbildung 11.1 (b). Die Anpassung der RovA-Konzentration an den Gleichgewichtszustand benötigt Zeit ( $t$ ). Die Zeit der Anpassung an eine niedrige Konzentration und die Zeit der Anpassung an höhere RovA-Konzentrationen unterscheiden sich (Abbildung 11.2).

In Abbildung 11.2 wird die Reaktionszeit der RovA-Anpassung von 25°C oder 37°C zu den Gleichgewichtskonzentrationen der jeweiligen Temperaturen in [°C] auf der  $x$ -Achse dargestellt. Die Parameter sind der Tabelle 10.4 entnommen. Die Gleichgewichtskonzentration  $r_g(T)$  ist abhängig von der Temperatur, jedoch nicht von der Zeit.

**Definition 13.** Eine RovA-Konzentration  $r(t)$  gilt an die Gleichgewichtskonzentration  $r_g$  **angepasst**, wenn

$$r(t) \in [0.95 \cdot r_g, 1.05 \cdot r_g]$$

ist.

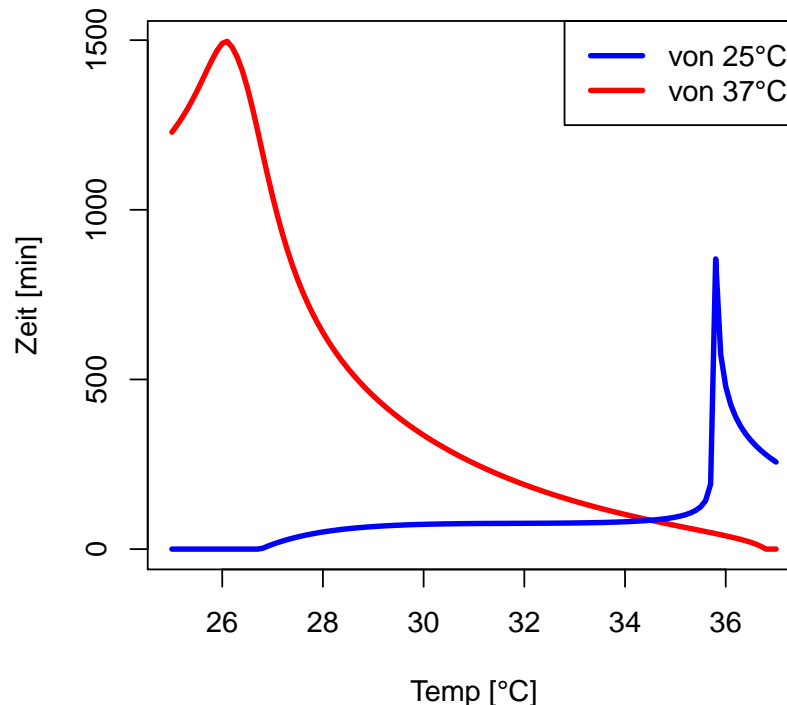
Die Reaktionszeit  $t_g$  wird in Minuten angegeben.

**Definition 14.** Gegeben sind eine Startzeit von  $t_0$  mit einer dazugehörigen RovA-Startkonzentration von  $r_0$ , also  $r(t_0) = r_0$ , und dem angepassten Zustand an die Gleichgewichtskonzentration von RovA  $\bar{r}_g \in [0.95 \cdot r_g, 1.05 \cdot r_g]$  mit  $r(t_g) = \bar{r}_g$ . So ist hier die **Reaktionszeit** als

$$t_r = t_g - t_0$$

definiert.

Die Anpassung von niedriger Temperatur (25° C) zu geringfügig höheren Temperaturen, bis 10° C mehr, geht schnell vonstatten. Erst wenn mindestens eine Temperatur von 36° C erreicht wird, benötigt die Anpassung wesentlich länger (von unter 100 min auf über 500 min). Dies lässt darauf schließen, dass genau bei dieser Temperatur ein Zustandswechsel stattfindet. Hat die Anpassung eine große Konzentrationsänderung zur Folge, benötigt ein System länger, um sich anzupassen.



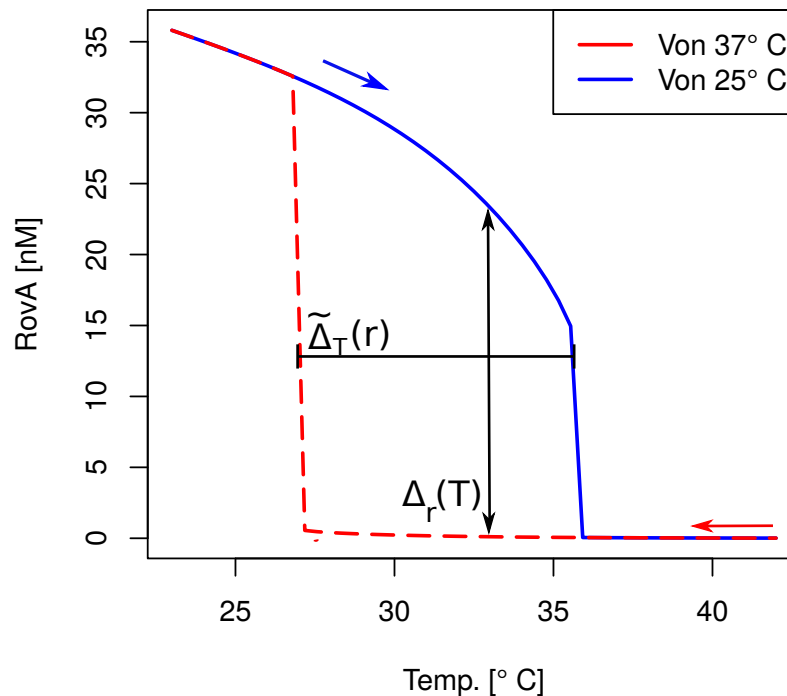
**Abbildung 11.2.: Reaktionszeit der RovA-Anpassung von 37° C und 25° C ausgehend.** Es wird die benötigte Zeit der Anpassung in Minuten dargestellt. Die Umgebungstemperatur (in [° C]), an die sich angepasst werden soll wird auf der x-Achse dargestellt. In rot wird die Anpassungsdauer von 37° C, in blau von 25° C, ausgehend dargestellt. Als angepasst gilt eine Konzentration, wenn mindestens 95 % des Gleichgewichtszustandes vorliegt.

Ist ein System in höheren Temperaturen (37° C) und wechselt es zu niedrigeren Temperaturen, so benötigt es im Gegensatz zu dem Wechsel von niedrigen Temperaturen auf höhere mehr Zeit, um sich anzupassen. Die Anpassung des Gleichgewichtszustandes bei einem Wechsel von 37° C auf unter 26° C benötigt besonders viel Zeit. Auch hier lässt sich ebenfalls auf einen Zustandswechsel schließen. Die beiden lokalen Extrema der Reaktionszeit bei 36° C und 26° C deuten darauf hin, dass zwischen diesen Temperaturen ein bistabiles Verhalten vorliegt. Die größere Reaktionszeit der Anpassung von hohen Temperaturen auf niedrigere Temperaturen zeigt, dass das System anstrebt, eher in niedrigeren Konzentrationen von den höheren Tem-

peraturen zu verweilen. Ein Bakterium, das in höheren Temperaturen lebt, braucht länger, um sich auf niedrige Temperaturen einzustellen als anders herum.

Bei niedrigen Temperaturen liegt eine höhere Konzentration von RovA vor als bei hohen Temperaturen (siehe Abbildung 11.3). Der Aufbau von RovA benötigt wesentlich mehr Zeit als der Abbau. Der verzögerte Aufbau liegt an der sehr niedrigen basalen Expressionsrate von RovA. Ist eine Zelle einmal inaktiv in der RovA-Produktion, benötigt diese einen großen Temperaturwechsel, um die RovA-Produktion zu aktivieren. Doch ist eine RovA-Konzentration vorhanden, wird die RovA-Produktion durch die Autoregulation aktiv gehalten. Die Temperatur muss hoch genug sein, um eine vollständige Aktivierung der Lon-Protease zu gewährleisten und somit den Abbau von RovA im Verhältnis zur Produktion überwiegen zu lassen.

Die Gleichgewichtszustände der RovA-Konzentration hängen direkt von der Temperatur ab. Durch numerische Berechnung der Gleichgewichtszustände kann gezeigt werden, dass nicht nur die Temperaturabhängigkeit in dem momentanen Zustand liegt, sondern ebenfalls in der Temperatur und somit der RovA-Konzentration des vorherigen Zustands. Es liegt eine Hysterese vor. In Abbildung 11.3 werden die Gleichgewichtszustände abhängig von der Temperatur berechnet. Es werden zwei unterschiedliche Startwerte angenommen, die je eine RovA-Gleichgewichtskonzentration  $r_g(T)$  bei  $25^\circ\text{C}$  bzw.  $37^\circ\text{C}$  widerspiegeln,  $r_g(25^\circ\text{C})$  und  $r_g(37^\circ\text{C})$ . Bei Temperaturen unter  $26^\circ\text{C}$  und Temperaturen über  $36^\circ\text{C}$  gibt es unabhängig von den Startwerten nur einen stabilen Zustand. Hier ist das System monostabil. Zwischen  $26^\circ\text{C}$  und  $36^\circ\text{C}$  gibt es jedoch zwei stabile Zustände. Der Unterschied zwischen diesen beiden stabilen Zuständen wird mit  $\Delta_r(T)$  bezeichnet mit der Einheit nM. Der Temperaturbereich der Bistabilität wird mit  $\tilde{\Delta}_T(r)$  bezeichnet und kann in  $^{\circ}\text{C}$  angegeben werden.



**Abbildung 11.3.: Gleichgewichtszustand der RovA-Konzentration abhängig von der Starttemperatur.** Die Temperatur auf der  $x$ -Achse wird in °C angegeben, die RovA-Konzentration auf der  $y$ -Achse in nM. Parameter wurden der Tabelle 10.4 entnommen. In rot wird der Gleichgewichtszustand der RovA-Konzentration von 37°C, in blau von 25°C, ausgehend dargestellt. Zwischen ca 26°C und 36°C ist ein bistabiles Verhalten der RovA-Konzentration zu erkennen. Der Unterschied zwischen beiden stabilen Zuständen wird mit  $\Delta_r(T)$ , der Temperaturbereich der Bistabilität wird mit  $\tilde{\Delta}_T(r)$  bezeichnet.

Das Differentialgleichungssystem verbleibt auch bei kleinen Temperaturanpassungen mit seiner RovA-Konzentration in einem Zustand. Findet ein Wechsel des Zustandes statt, so benötigt dieser Wechsel einen größeren Temperaturreiz und eine gewisse Zeit der Anpassung (siehe Abbildung 11.2).

Der Temperaturbereich der Bistabilität kann direkt aus dem Modell berechnet werden. Der Temperaturbereich der Bistabilität und dessen Größe  $\Delta_r(T)$

kann auch für variierende Parameter dargestellt werden.  $\Delta_T(T)$  ist der Unterschied der Konzentrationsgröße der beiden stabilen Zustände und wird in nM angegeben. Ein Fokus dieser Arbeit liegt auf der Robustheit des Temperaturbereichs der Bistabilität  $\tilde{\Delta}_T(r)$  bei Veränderung der basalen Expressionsrate  $\alpha_0$ , der maximalen Expressionsrate  $\alpha$ , der Degradationsrate  $\delta(T)$  und der Einfluss des reprimierenden Promotors.

### 11.2.1. Einfluss der basalen Expressionsrate

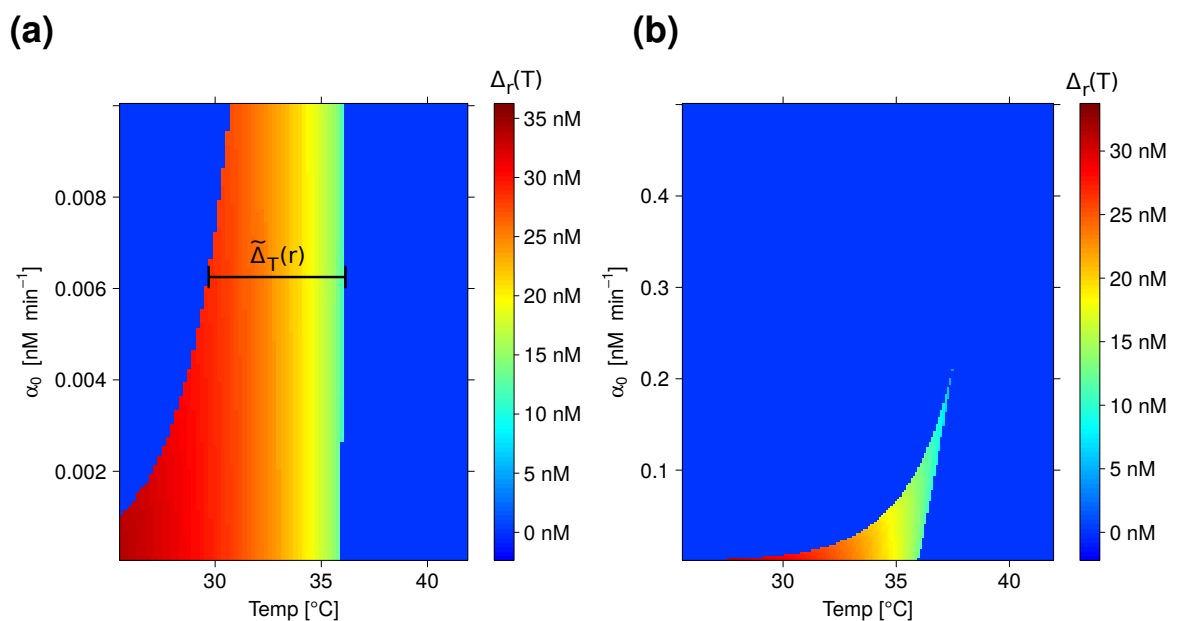
Die basale Expressionsrate  $\alpha_0$  ist ein wichtiger Faktor innerhalb des RovA-Modells. Sie zeigt, wie stark RovA produziert wird, wenn der RovA-Promotor nicht angeschaltet ist, die Produktion also eigentlich nicht induziert ist. Zu klären ist, wie stark der Einfluss der basalen Expressionsrate  $\alpha_0$  auf die Bistabilität ist.

Bei einer sehr großen basalen Expressionsrate kommt es nur noch zu monostabilem Verhalten. Die basale Expression von RovA aktiviert dauerhafte den Promotor und es kommt zu einer Stabilität der Konzentration durch die Autoregulation der Produktion. Es würde dann bei jeder Temperatur der Zustand der höheren RovA-Konzentration vorliegen. Ist keine basale Expressionsrate vorhanden, kann das System nicht mehr produzieren.

Der Gleichgewichtszustand wird in Abbildung 11.4 für RovA-Konzentrationen  $r_g$  von 25° C und von 37° C ausgehend berechnet. Die Zieltemperatur  $T$  und die basale Expressionsrate  $\alpha_0$  werden variiert. Die anderen Parameter wurden aus Tabelle 10.4 entnommen. In Abbildung (a) befindet sich die basale Expressionsrate innerhalb des Intervalls ]0 nM/min, 0.01 nM/min]. Auf dem Intervall ]0 nM/min, 0.5 nM/min] liegt die basale Expressionsrate in



Abbildung (b). Der Farbcode gibt die Größe von  $\Delta_r(T)$  an, also den Unterschied zwischen den stabilen Zuständen der RovA-Konzentrationen. Bei dunkelblauer Farbe existiert kein Unterschied zwischen beiden Zuständen ( $\Delta_r(T) = 0$ ), Monostabilität liegt vor. Bei allen anderen Farben sind zwei unterschiedliche stabile Zustände vorhanden.



**Abbildung 11.4.: Phasendiagramm des bistabilen Verhaltens des RovA-Systems abhängig von der Temperatur und der basalen Expressionsrate  $\alpha_0$ .** Durch numerische Berechnung wird der Gleichgewichtszustand der RovA-Konzentration zu variablen Temperaturen berechnet, die sich durch die Startwerte unterscheiden. Hierbei werden Startwerte genutzt, die einer RovA-Konzentration bei 25°C und bei 37°C entsprechen. Die fehlenden Parameter werden der Tabelle 10.4 entnommen. Auf der  $x$ -Achse wird die Zieltemperatur in [°C] und auf der  $y$ -Achse die basale Expressionsrate  $\alpha_0$  in nM/min variiert. Die Skalierung von  $\alpha_0$  in Abbildung (a) [0 nM/min, 0.01 nM/min] und in Abbildung (b) [0 nM/min, 0.5 nM/min]. Der Farbcode zeigt an, wie groß der Unterschied zwischen den beiden Gleichgewichtskonzentrationen von RovA ist. Dunkelblau bedeutet, dass der Unterschied bei Null nM, daher liegt dort ein monostabiles Verhalten vor. Dunkelrot bedeutet ein besonders großen Unterschied zwischen beiden stabilen Zuständen, somit ein besonders großes  $\Delta_r(T)$ . Der Temperaturbereich der Bistabilität  $\tilde{\Delta}_T(r)$  hängt von der RovA-Konzentration und somit direkt von den Parametern ab.

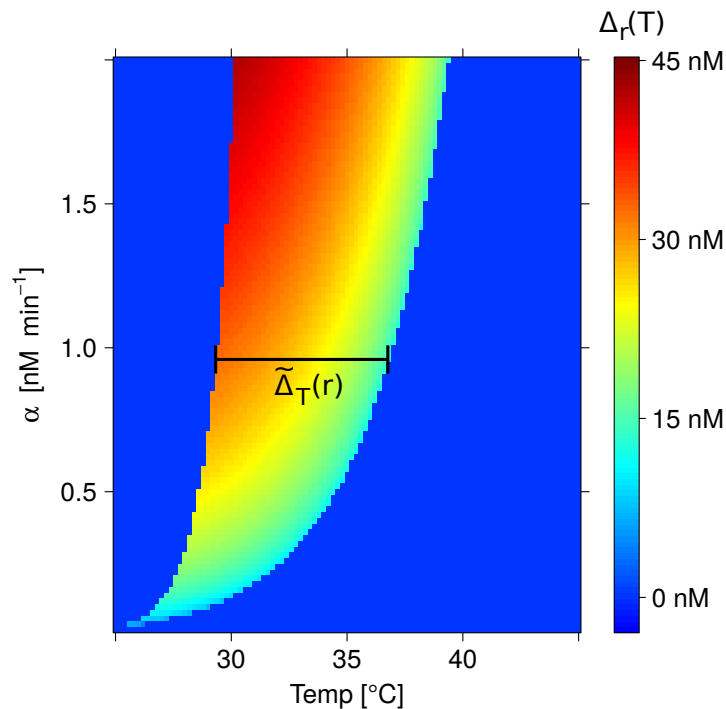
Es lässt sich in Abbildung 11.4 (a) erkennen, dass wenn die basale Expressionsrate gegen 0 nM/min geht, der bistabile Temperaturbereich immer größer wird und auch niedrigen Temperaturen ein bistabiles Verhalten aufweisen. An Abbildung 11.4 (b) zeigt sich, dass sobald die basale Expressionsrate größer ist als 0.2 nM/min, gar keine Bistabilität mehr vorliegt.

Um einen Temperaturbereich des bistabilen Verhaltens  $\tilde{\Delta}_T(r)$  zu erreichen, der exakt auf dem Temperaturbereich liegt, auf dem die Bistabilität von RovA in *Y. pseudotuberculosis* vorliegt ([26° C, 37° C]), muss die basale Expressionsrate exakt bestimmt werden. In Kapitel 10.4 wird diese berechnet mit  $\alpha_0 = 0.0018724$  nM/min und liegt damit auf dem Temperaturbereich, bei dem es das bistabile Verhalten von RovA von *Y. pseudotuberculosis* in der Natur widerspiegelt. Das Bakterium muss eine sehr exakte und robuste basale Expressionsrate von RovA besitzen, sonst könnte es nicht immer das gleiche bistabile Verhalten zeigen. Daher muss die basale Expressionsrate robust gegen äußere Einflüsse sein.

### 11.2.2. Einfluss der maximalen Expressionsrate

Bei größerem  $\alpha$  steigert sich der Einfluss der Autoregulation. Um von hoher RovA-Konzentration auf eine niedrige zu wechseln, wird bei einer größeren maximalen Expressionsrate ein stärkerer Einfluss der Lon-Protease benötigt, damit die Autoregulation kompensiert werden kann. Liegt eine niedrige RovA-Konzentration vor, kann bei stärkerer Autoregulation die RovA-Produktion schneller starten. Mit einer höheren maximalen Expressionsrate  $\alpha$  kann sich der bistabile Temperaturbereich vergrößern. Er verlagert sich dabei immer weiter in die höheren Temperaturbereiche. Der Tempe-

raturbereich der Bistabilität  $\tilde{\Delta}_T(r)$  verschiebt sich auf höhere Temperaturen.



**Abbildung 11.5.: Phasendiagramm des bistabilen Verhaltens des RovA-Systems abhängig von der Temperatur und der maximalen Expressionsrate  $\alpha$ .**

Der Gleichgewichtszustand der RovA-Konzentration wird von 25°C und von 37° ausgehend berechnet, daher für unterschiedliche Startwerte. Die fehlenden Parameter werden der Tabelle 10.4 entnommen. Auf der x-Achse ist die Zieltemperatur und auf der y-Achse die maximale Expressionsrate dargestellt. Der Unterschied zwischen beiden stabilen Zuständen wird durch die rechtsstehende Farbskala gezeigt und beträgt maximal 45 nM.

In Abbildung 11.5 wird mittels eines Phasendiagramms gezeigt, welchen Einfluss die maximale Expressionsrate  $\alpha$  auf den Temperaturbereich der Bistabilität  $\tilde{\Delta}_T(r)$  hat. Je kleiner  $\alpha$  ist, um so kleiner ist der bistabile Bereich. Die maximale Expressionsrate  $\alpha$  wird auf dem Intervall ]0 nM/min, 2.0 nM/min] variiert. Die anderen Parameter sind aus Tabelle 10.4 entnommen. Das bistabile System hat eine festgelegte maximale Expressionsrate. Das bistabile

Verhalten des Systems ist jedoch bei Veränderungen von  $\alpha$  nicht so anfällig wie bei Veränderung von  $\alpha_0$ .

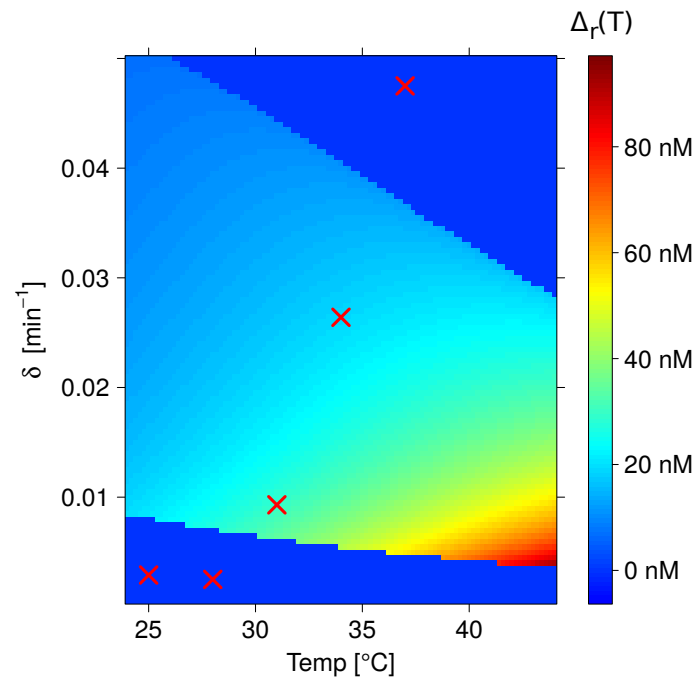
Die experimentellen Daten ergaben ein  $\alpha$  von 0.7110374 nM/min (Tabelle 10.4). Dieses angegebene  $\alpha$  befindet sich in exakt den bistabilen Temperaturbereich, den RovA in *Y. pseudotuberculosis* besitzt. Bei kleinen Veränderungen ( $\pm 10\%$ ) von  $\alpha$  schwankt der bistabile Temperaturbereich nicht so stark wie bei kleinen Veränderung von  $\alpha_0$ . Kleine Veränderungen der maximalen Expressionsrate  $\alpha$  haben keinen großen Einfluss auf die Bistabilität. Dies ist für das Bakterium von Vorteil, da  $\alpha$  bei Änderungen der äußeren Umgebung sich ebenfalls ändern könnte.

### 11.2.3. Einfluss der Degradation

Die Degradationsrate  $\delta(T)$  ist im Gegensatz zu der basalen Expressionsrate  $\alpha_0$  und der maximalen Expressionsrate  $\alpha$  direkt abhängig von der Temperatur. Der Einfluss von

$$\delta(T) = 5.2 \cdot 10^{-6} e^{0.2469T}$$

auf die Bistabilität und die Frage, ob die Temperaturabhängigkeit der Abbaurate eine notwendige Bedingung für diese ist, ist daher besonders interessant. Diese experimentell bestimmten Werte der Abbaurate sollen innerhalb eines Phasendiagramms wie in Kapitel 11.2.1 und 11.2.2 dargestellt werden.



**Abbildung 11.6.: Phasendiagramm des bistabilen Verhaltens des RovA-Systems abhängig von der Temperatur und der Abbaurate  $\delta(T)$ .**

Numerisch wird der Gleichgewichtszustand der RovA-Konzentration für unterschiedliche Startwerte berechnet. Eine RovA-Konzentration zeigen den Unterschied der beiden stabilen Zustände an. Die fehlenden Parameter werden der Tabelle 10.4 entnommen. Auf der  $x$ -Achse wird die Zieltemperatur und auf der  $y$ -Achse die Abbaurate dargestellt. Der Unterschied zwischen beiden stabilen Zuständen wird durch die rechtsstehende Farbskala gezeigt und beträgt maximal 90 nM. Die roten Kreuze zeigen die im Experiment gemessenen Degradationsraten entsprechend zu den Temperaturen.

Das Phasendiagramm in Abbildung 11.6 zeigt das bistabile Verhalten des Gleichgewichtszustandes von RovA  $r_g$ , abhängig von den Anfangstemperaturen, der Temperatur selbst und einer Abbaurate von  $\delta \in ]0, 0.5]$  1/min. So kann man in das Phasendiagramm die fünf gemessenen Abbauwerte abhängig von Temperatur eintragen und erkennen, welchen Einfluss diese auf die Bistabilität hat.

Die Bistatilität des RovA-Systems ist direkt abhängig von der Degradation  $\delta(T)$ . Das temperaturabhängige bistabile Verhalten, das das RovA-System

zeigt, ist nur durch die temperaturabhängige Degradationsrate möglich. Würde die Degradationsrate nicht temperaturabhängig sein, so würde zumeist gar keine Bistabilität oder ausschließlich Bistabilität auftreten.

Durch die gemessenen Daten von  $\delta(T)$  und die daraus bestimmte exponentielle Funktion, abhängig von der Temperatur mit den in Tabelle 10.4 beschriebenen Parametern, bestätigt sich der bistabile Bereich, der in der Natur bei dem RovA-System vorkommt. Nicht nur die Größe von  $\delta(T)$  sondern auch die Abhängigkeit von der Temperatur ist eine notwendige Bedingung für diesen Bereich der Bistabilität.

### 11.3. Erstellen der *in silico*-Mutante $\Delta iRovA$

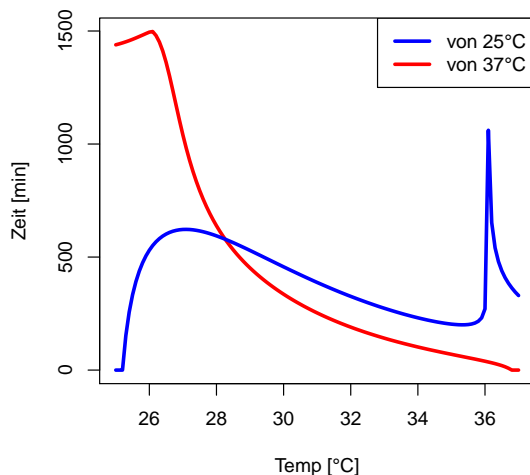
Das bistabile Verhalten des RovA-Systems entsteht durch die positive FBL (Angeli et al., 2004). Um den Einfluss der negativen FBL (NFL) auf das RovA-System zu untersuchen, wird eine *In silico*-Mutante erstellt, bei der das RovA-System keine negative FBL besitzt. Die Mutante wird mit  $\Delta iRovA$  bezeichnet. In den vorherigen Kapiteln konnte gezeigt werden, dass die RovA-Differentialgleichung mit den berechneten Parametern das wahre Verhalten von RovA ziemlich präzise widerspiegelt. Es kann von dem obigen Modell ausgegangen werden.

Die Produktion von RovA innerhalb der Mutante  $\Delta iRovA$  soll unabhängig von der NFL sein. Die NFL geht mit  $(k_r(T)^{h_r})/(k_r(T)^{h_r} + r^{h_r})$  in die RovA-Produktion ein. Dieser Term wird nun innerhalb der Differentialgleichung entfernt. Nun besitzt die RovA-Produktion folgende Differentialgleichung:

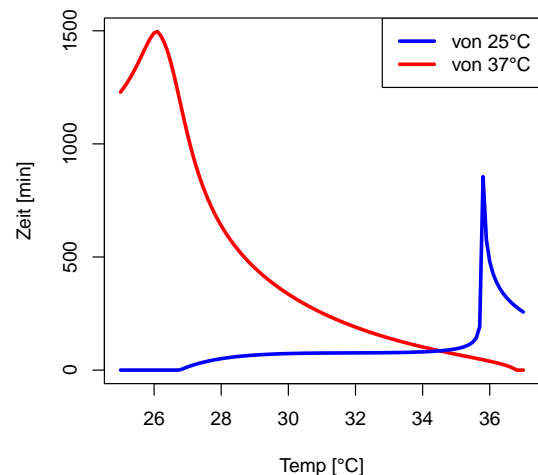
$$\dot{r}(t) = \alpha_0 + \alpha \cdot \frac{r^{h_a}}{k_a(T)^{h_a} + r^{h_a}} - \delta(T) \cdot r. \quad (11.1)$$

Die enthaltenen Parameter in Gleichung (11.1) entsprechen denen aus der Tabelle 10.4. Durch die neu entstandene Differentialgleichung kann ein neues Verhalten des RovA-Systems entstehen. Das veränderte Verhalten zeigt, welchen Einfluss der reprimierende Promotor und dadurch die NFL auf das gesamte System hat.

(a)



(b)



**Abbildung 11.7.: Reaktionszeit der RovA-Anpassung von höherer und niedriger Temperatur des Mutationssystems.**

Es wird die benötigte Zeit der Anpassung in Minuten dargestellt. Die Umgebung, an die sich angepasst werden soll, wird auf der  $x$ -Achse dargestellt. Die Parameter sind der Tabelle 10.4 entnommen. In blau wird die Anpassungsdauer von einer Startkonzentration ausgegangen, die bei 25°C vorliegt (höhere Konzentration); in rot von einer Startkonzentration die bei 37°C vorliegt. Als angepasst gilt eine Konzentration, wenn mindestens 95 % des Gleichgewichtszustandes vorliegt. In (b) dargestellt die Reaktionszeit des Wildtyps, in (a) die der *in silico*-Mutante.

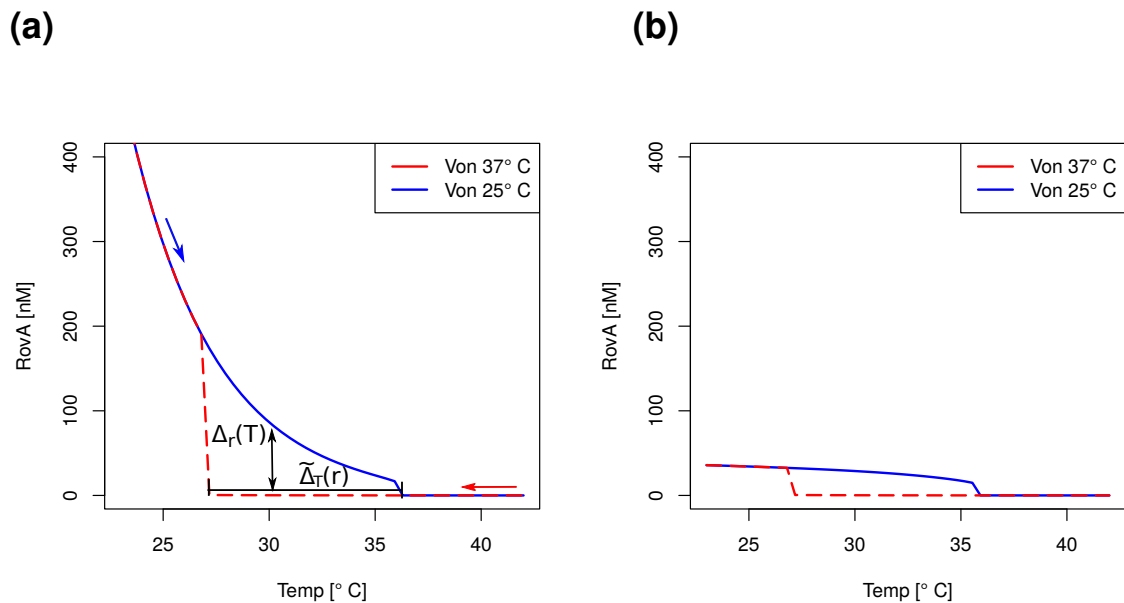
In Abbildung 11.7 werden Reaktionszeiten dargestellt, die ein bei 25° C,

bzw. ein bei 37° C lebendes Bakterium benötigt, um sich auf einen Temperaturwechsel zu höheren, bzw. niedrigere Temperaturen anzupassen. Die Anpassung an den Gleichgewichtszustand  $r_g$  wird in Definition 13 dargestellt. In Definition 14 wird die Reaktionszeit  $t_r$  definiert. Die benötigte Zeit wird in den Abbildungen 11.7 (a) + (b) auf der y-Achse dargestellt, die Endtemperatur nach dem Wechsel wird auf der x-Achse aufgetragen.

Man kann an den Reaktionszeiten nach einem Temperaturwechsel in Abbildung 11.7 erkennen, dass die  $\Delta$ iRovA-Mutante (a) deutlich andere Reaktionszeiten hat als der Wildtyp (b). In blau dargestellt wird von einem System bei 25° C ausgegangen. Hier liegt die Reaktionszeit der Mutante wesentlich höher als die Reaktionszeit des Wildtyps. In rot wird von einem System bei 37° C ausgegangen, bei der sich die Reaktionszeiten kaum unterscheiden. Die NFL hat daher besonderen Einfluss auf die Reaktionszeiten von niedrigen zu höheren Temperaturen. Wenn ein Bakterium von einer Umgebungstemperatur zu einer Wirtstemperatur wechselt, würde das mutierte Bakterium ohne NFL wesentlich länger brauchen sich anzupassen als der Wildtyp. Die schnelle Anpassung von *Y. pseudotuberculosis* an eine Wirtsumgebung ist notwendig für eine schnelle Kolonialisierung und Infektion. Dauert dies zu lange, kann möglicherweise das Immunsystem des Wirtes reagieren und die Bakterien abtöten.

In Abbildung 11.8 wird sowohl der Gleichgewichtszustand der RovA-Konzentration von der *in silico*-Mutante als auch der des Wildtyps gezeigt. Die Gleichgewichtskonzentrationen von beiden zeigen deutlich Größenunterschiede. Die Konzentration von RovA der  $\Delta$ iRovA-Mutante ist bei niedrigen Temperaturen wesentlich höher als beim Wildtyp. Bei hohen Temperaturen sind die Gleichgewichtskonzentrationen sowohl bei der Mutante, als auch beim Wildtyp fast bei 0 nM.





**Abbildung 11.8.: Gleichgewichtszustand der RovA-Konzentration abhängig von der Starttemperatur der  $\Delta i$ RovA-Mutante und des Wildtyps.**

Die Temperatur auf der  $x$ -Achse wird in  $^{\circ}\text{C}$  angegeben, die RovA-Konzentration auf der  $y$ -Achse in nM. Die Parameter sind der Tabelle 10.4 entnommen. In blau wird der Gleichgewichtszustand der Konzentration von RovA gezeigt mit einem Startwert, der der Konzentration von RovA bei  $25^{\circ}\text{C}$  entspricht. Der Gleichgewichtszustand der RovA-Konzentration von  $37^{\circ}\text{C}$  ausgehend wird in rot dargestellt. In Abbildung (a) wird der Gleichgewichtszustand der RovA-Konzentration der  $\Delta i$ RovA-Mutante dargestellt, in (b) zum Vergleich wie in Abbildung 11.3 der Gleichgewichtszustand der RovA-Konzentration des Wildtyps. Für einen besseren Vergleich wurde in Abbildung (b) die  $y$ -Achsen-Skalierung angepasst.

Die Temperaturbereiche der Bistabilität  $\tilde{\Delta}_T(r)$  sind bei beiden Modellen gleich. An Abbildung 11.8 (a),  $\Delta i$ RovA-Mutante, kann man erkennen, dass bei niedrigeren Temperaturen die RovA-Konzentration schnell größer wird im Gegensatz zum Wildtyp (Abbildung 11.8 (b)). Durch diese größere Konzentration benötigt das System auch mehr Zeit für die Anpassung (Abbildung 11.7), da die Konzentrationsunterschiede der verschiedenen Gleichgewichtszustände abhängig von der Temperatur größere Unterschiede aufweisen als bei dem Wildtyp. Der Konzentrationsunterschied zwischen beiden

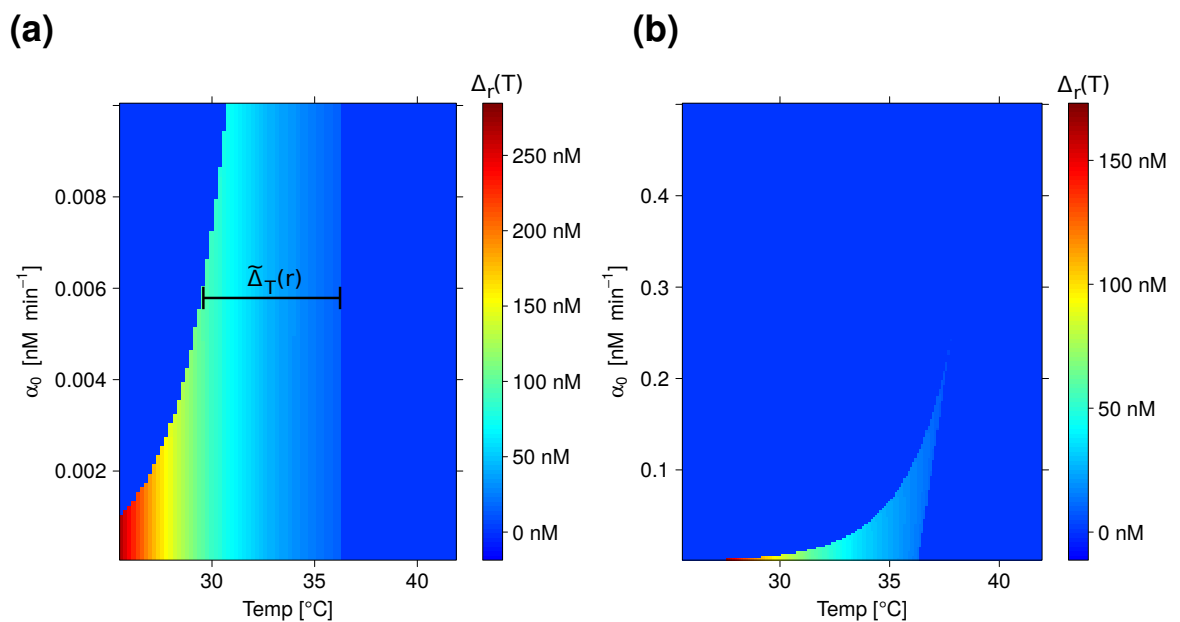
stabilen Zuständen  $\Delta_r(T)$  ist dadurch ebenfalls größer. Sowohl die Zeit der Anpassung ist durch die NFL um ein vielfaches verbessert, als auch dass der Aufbau der hohen Konzentration von RovA bei niedrigen Temperaturen, was dem Bakterium sehr viel Energie kostet. Mit der NFL kann das Bakterium diese energetischen Kosten einsparen.

## 11.4. Analyse der Bistabilität der *in silico*-Mutante $\Delta iRovA$

Um das bistabile Verhalten von RovA in der Mutante näher zu untersuchen, wurden für  $\Delta iRovA$  ebenfalls Phasendiagramme mit variabler basaler Expressionsrate, maximaler Expressionsrate und Degradationsrate erstellt.

### 11.4.1. Einfluss der basalen Expressionsrate

Das Phasendiagramm mit variabler basale Expressionsrate  $\alpha_0$  innerhalb des Wildtyps (Abbildung 11.4) zeigt, dass keine großen Veränderungen der basalen Expressionsrate möglich sind, wenn das System weiterhin auf dem gleichen Temperaturbereich Bistabilität zeigen soll. Das System ist also nicht robust gegenüber Schwankungen der basalen Expressionsrate.



**Abbildung 11.9.: Phasendiagramm des bistabilen Verhaltens des RovA-Mutationssystems abhängig von der Temperatur und der basalen Expressionsrate  $\alpha_0$ .**

Numerisch wird der Gleichgewichtszustand von der RovA-Konzentration für unterschiedliche Startwerte berechnet. Eine RovA-Konzentration entspricht der Konzentration bei Temperaturen um die 37° C und bei 25° C. Die fehlenden Parameter sind der Tabelle 10.4 entnommen. Auf der x-Achse wird die Zieltemperatur in [° C] und auf der y-Achse die basale Expressionsrate  $\alpha_0$  in nM/min variiert. Die Skalierung von  $\alpha_0$  in Abbildung (a) [0 nM/min, 0.01 nM/min] und in Abbildung (b) [0 nM/min, 0.5 nM/min]. Der Konzentrationsunterschied zwischen beiden stabilen Zuständen wird mit  $\Delta_r(T)$  beschrieben. Der Farbcode zeigt an, wie hoch  $\Delta_r(T)$  ist. Bei dunkelblau liegt der Unterschied bei Null ( $\Delta_r(T) = 0$ ), daher liegt dort ein monostabiles Verhalten vor. Dunkelrot bedeutet ein besonders großen Unterschied zwischen beiden stabilen Zuständen. Das Temperaturintervall auf dem die Bistabilität liegt, wird mit  $\tilde{\Delta}_T(r)$  bezeichnet.

In Abbildung 11.9 (a) und (b) wird das bistabile Verhalten der Gleichgewichtszustände von RovA der *in silico*-Mutante gezeigt. In Abbildung 11.9 (a) wird  $\alpha_0$  auf dem Intervall [0 nM/min, 0.01 nM/min] und bei (b) auf dem Intervall [0 nM/min, 0.5 nM/min] variiert.

Vergleicht man Abbildung 11.4 mit Abbildung 11.9, so kann man erkennen,

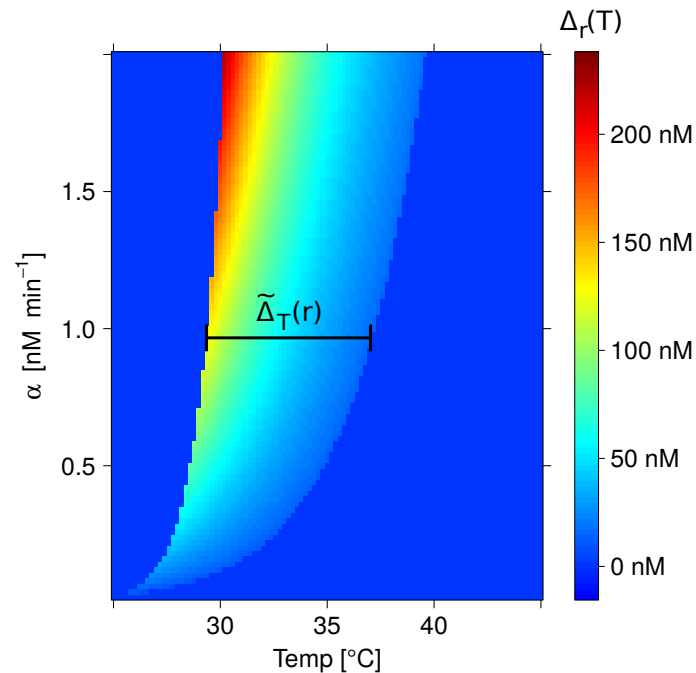
dass die Phasen der Bistabilität bei beiden, Wildtyp und Mutante, nahezu gleich bleiben. Der Unterschied zwischen den beiden Modellen zeigt auch hier, dass die Mutante einen Unterschied zum Wildtyp nur in dem Konzentrationsunterschied der RovA-Gleichgewichtszustände zeigt.

### 11.4.2. Einfluss der maximalen Expressionsrate

Der Einfluss der maximalen Expressionsrate  $\alpha$  wurde ebenfalls für die  $\Delta iRovA$ -Mutante untersucht und in einem Phasendiagramm in Abbildung 11.10 dargestellt. Innerhalb des RovA-Systems des Wildtyps hat  $\alpha$  gezeigt, dass sich bei größerem  $\alpha$  der Temperaturbereich der Bistabilität vergrößert und gleichzeitig zu höheren Temperaturen verschiebt (siehe Abbildung 11.5). Zu untersuchen ist, ob die maximale Expressionsrate bei der  $\Delta iRovA$ -Mutante den gleichen Einfluss hat.

Innerhalb der  $\Delta iRovA$ -Mutante zeigt  $\alpha$  den gleichen Einfluss auf die temperaturabhängige Bistabilität wie in dem Wildtyp. Bei immer kleiner werdender maximalen Expressionsrate  $\alpha$  kommt es zu einem immer kleineren Temperaturintervall  $\tilde{\Delta}_T(r)$ , auf dem das System bistabil ist.

Doch auch hier gibt es einen erheblicher Unterschied auf der Konzentrationsskala  $\Delta_r(T)$ , welcher den Unterschied zwischen den beiden stabilen Zuständen zeigt. In Abbildung 11.5 kann man erkennen, dass höchstens ein Konzentrationsunterschied zwischen den beiden stabilen Zuständen der RovA-Konzentration von 45 nM vorliegt. Innerhalb der Mutante steigt dieser Unterschied auf 200 nM.



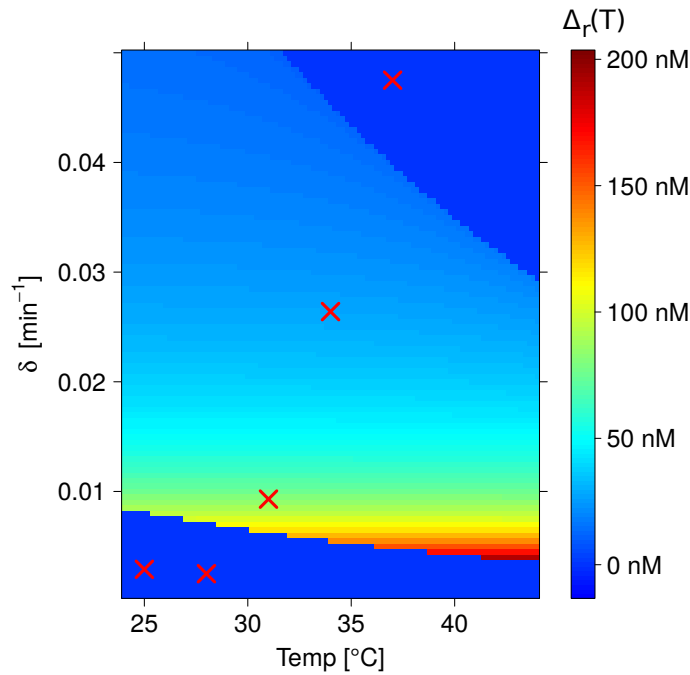
**Abbildung 11.10.: Phasendiagramm des bistabilen Verhaltens des RovA-Mutationssystems abhängig von der Temperatur und der maximalen Expressionsrate  $\alpha$ .**

Durch numerische Berechnung ist sowohl der Gleichgewichtszustand von der RovA-Konzentration der  $\Delta iRovA$ -Mutante von niedrigen, wie von hohen Temperaturen berechnet, also für unterschiedliche Startwerte. Der Konzentrationsunterschied zwischen beiden stabilen Zuständen wird mit  $\Delta_r(T)$  beschrieben, welche durch die Farbskalierung dargestellt wird. Hier ist auf der  $x$ -Achse die Zieltemperatur und auf der  $y$ -Achse die maximale Expressionsrate variiert. Die fehlenden Parameter sind der Tabelle 10.4 entnommen. Das Temperaturintervall auf dem die Bistabilität liegt wird mit  $\tilde{\Delta}_T(r)$  bezeichnet.

### 11.4.3. Einfluss der Degradation

Der Einfluss der Degradation  $\delta(T)$  auf das bistabile Verhalten abhängig von der Temperatur innerhalb des Wildtyps zeigt, dass in diesem System die temperaturabhängige Degradation  $\delta$  notwendig ist, damit das RovA-System innerhalb der angegebenen Temperatur bistabil ist (siehe Abbildung 11.6).

Auch für das System der  $\Delta iRovA$ -Mutante wird der Einfluss der Degradation auf die Bistabilität innerhalb eines Phasendiagramms in Abbildung 11.11 dargestellt. Die Messdaten aus Tabelle 10.2 sollen ebenfalls in das Phasendiagramm eingetragen werden.



**Abbildung 11.11.: Phasendiagramm des bistabilen Verhaltens des RovA-Mutationsystems abhängig von der Temperatur und der Abbaurate  $\delta(T)$ .**

Numerisch wird der Gleichgewichtszustand von der RovA-Konzentration für unterschiedliche Startwerte berechnet. Eine RovA-Konzentration entspricht der Konzentration bei Temperaturen bei 37° C und bei 25° C. Hier wird auf der  $x$ -Achse die Zieltemperatur und auf der  $y$ -Achse die Abbaurrate variiert. Die roten Kreuze zeigen die gemessenen Degradationsraten passend zu den Temperaturen. Die fehlenden Parameter sind der Tabelle 10.4 entnommen.

Genau wie bei den Expressionsraten ändert sich der Einfluss auf die Bistabilität von  $\delta$  innerhalb der Mutante im Gegensatz zum Wildtyp nicht. Der einzige Unterschied besteht innerhalb des Konzentrationsunterschiedes  $\Delta_r(T)$  zwischen den beiden stabilen Zuständen der RovA-Konzentration.

Es ergibt sich also deutlich, dass die Mutation des RovA-Systems keinen Einfluss auf den Temperaturbereich der Bistabilität  $\tilde{\Delta}_T(r)$  hat, jedoch einen sehr großen Einfluss auf die RovA-Konzentration und dadurch auch auf die Reaktionszeit bei niedrigeren Temperaturen.

Auf die Konzentration oder das Verhalten des RovA-Systems bei höheren Temperaturen ( $\geq 36^\circ \text{ C}$ ) hat die reprimierende Bindestelle, also die NFL, überhaupt keinen Einfluss.

#### 11.4.4. Potential

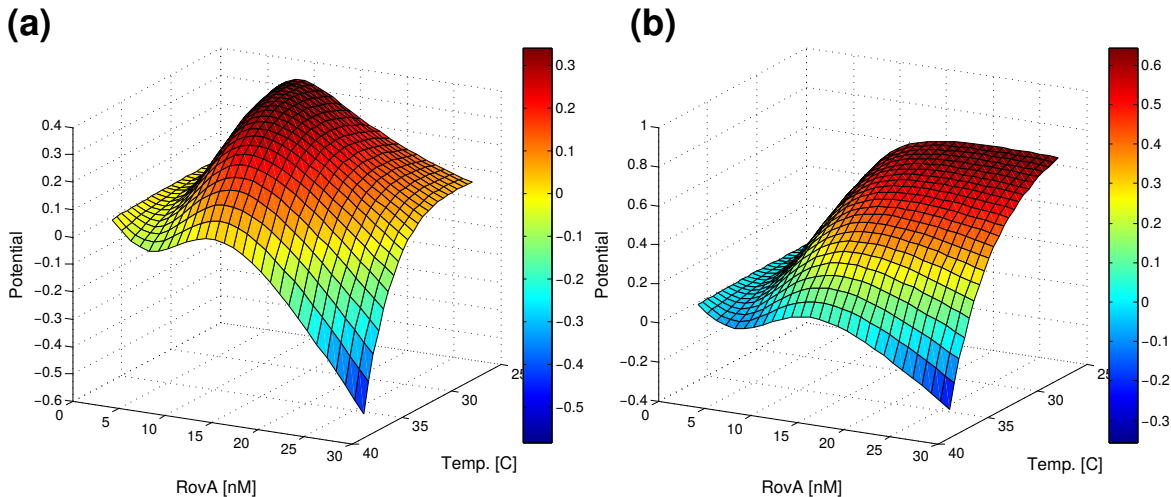
Innerhalb des RovA-Systems ist das Potential  $U(r)$  (Definition 11) abhängig von der RovA-Konzentration und der Temperatur und wird in Abbildung 11.12 in einer 3D-Landschaft dargestellt. Es kann mittels der rechten Seite der Differentialgleichung der RovA-Konzentration (Gleichung (10.10)) berechnet werden.

$$\frac{dU(r)}{dr} = -f(r) \quad (11.2)$$

In Abbildung 11.12 (a) wird das Potential des RovA-System vom Wildtyp dargestellt. Ein Maximum existiert bei ca. 10 nM. Dieses symbolisiert den instabilen Zustand des RovA-Systems. Es existieren zwei lokale Minima. Eines ist deutlich bei fast 0 nM RovA zu erkennen, das andere liegt bei höheren RovA-Konzentrationen ca. 30 nM.

Das System hat zwei stabile Zustände, die abhängig von der Temperatur sind. Im Vergleich zu Abbildung 11.12 (a) wird in Abbildung 11.12 (b) das Potential des RovA-Systems in der  $\Delta i\text{RovA}$ -Mutante gezeigt. Hierbei lässt

sich erkennen, dass das Minimum bei niedrigen RovA-Konzentrationen genau wie bei dem Modell des Wildtyps liegt.



**Abbildung 11.12.: Skalarpotential des RovA-Modells abhängig von der Temperatur und der RovA-Konzentration.**

In (a) wird das Potential des Wildtypmodells, in (b) wird das Potential der  $\Delta iRovA$ -Mutante gezeigt. Die Parameterdaten sind der Tabelle 10.4 entnommen.

Der Unterschied liegt bei dem Potential bei höheren RovA-Konzentrationen. Man kann erkennen, dass das Minimum des Potentials nicht schon bei ca. 30nM liegt, sondern viel höher liegen muss (siehe Abbildung 11.8 (a) - bei ab 100nM). Man kann auch aus Abbildung 11.8 (a) und (b) erkennen, da die RovA-Konzentration einer Mutante im Gleichgewichtszustand bei wesentlich höherer Konzentration liegt als bei dem Wildtyp.

## 11.5. Modell und natürliches Verhalten von RovA im Vergleich

Durch den deterministischen Ansatz konnte ein Modell (Kapitel 10.5) erstellt werden, welches das natürliche Verhalten von RovA innerhalb von



*Y. pseudotuberculosis* widerspiegelt. Durch das Einfließen der biologischen Daten, die in Kapitel 10.2 bis 10.4 präsentiert wurden, konnte ein realistisches Modell produziert werden, das der Wirklichkeit sehr nahe kommt und somit ein mechanistisches Modell darstellen könnte. Genau bei dem bistabilen Temperaturbereich wurde das Temperaturwechsel-Experiment (Kapitel 10.4) durchgeführt. Das Modell zeigt der Wirklichkeit entsprechendes Temperaturverhalten, Bistabilität und Hysterese. Somit kann das Modell benutzt werden, um realitätsnahe Vorhersagen zu treffen und bessere Systemeinsichten zu bekommen. Das RovA-System innerhalb *Y. pseudotuberculosis* ist in der Biologie eines der ersten gefundenen Beispiele für eine Bistabilität abhängig von der Temperatur, wie es bis jetzt nur beim  $\lambda$ -Phagen gezeigt werden konnte.

Wie bereits in Kapitel 11.2 beschrieben, zeigt das Modell unterschiedliche Reaktionszeiten, je nach Ausgangstemperatur und Endtemperatur. Dieses Verhalten ist eine Grundvoraussetzung der Bistabilität in Bakterien. Der dadurch entstandene bistabile Bereich des Systems liegt circa auf dem Intervall [26°C, 36°C]. Das ist genau das Temperaturintervall, welches zwischen Umwelttemperatur (unter 26°C) und Wirtstemperatur (ab 36°C) liegt. Ein stabiler Bereich des RovA-Modells besitzt eine RovA-Konzentration von annähernd 0 nM und der zweite stabile Bereich liegt bei etwas über 30 nM. Durch dieses bistabile Verhalten ist das gesamte RovA-System robust gegen kurzzeitige Änderungen der Umwelt. Die Reaktionszeit bei einem Zustandsunterschied liegt zwischen 600 min und 1500 min, also zwischen 10 h und 25 h. Ist der Temperaturwechsel nur kurzzeitig, so wechselt das System nicht vollständig in den anderen Zustand. Erhöht sich die Temperatur auf einen Bereich, bei dem das System Bistabilität zeigt, so ändert sich das Verhalten nicht. Erst bei Temperaturwechsel, die darüber hinaus gehen verändert sich das Verhalten des Systems.

Das System wurde auf seine Robustheit bezüglich der Bistabilität unter-

sucht. Bei Veränderungen der Parameter kommt es zu einem anderen Verhalten in der Bistabilität. Hierbei zeigt sich, dass nur bei einer sehr kleinen basalen Expressionsrate  $\alpha_0$  von RovA ein bistabiles Verhalten überhaupt möglich ist. Ist die basale Expressionsrate zu groß, wird das System monostabil. Im Gegensatz zur Robustheit der Bistabilität auf Veränderung der basalen Expressionsrate ist die Robustheit der Bistabilität bei Veränderungen der maximalen Expressionsrate  $\alpha$  wesentlich größer. Dies erscheint sinnvoll, da in der Realität die maximale Expressionsrate durch weitere Umweltparameter, wie zum Beispiel pH-Wert, Sauerstoffgehalt oder Salzgehalt leicht schwanken kann.

Eine notwendige Bedingung für das gezeigte bistabile Verhalten innerhalb des RovA-Systems ist die in *Y. pseudotuberculosis* vorliegende temperaturabhängige Konformationsänderung von RovA (siehe Kapitel 9.1.2). Dadurch entstehen die temperaturabhängigen DNA-Bindekonstanten und die temperaturabhängige Abbaurate von RovA. Der Abbau von RovA wird innerhalb von *Y. pseudotuberculosis* mittels der Lon-Protease durchgeführt (siehe Kapitel 9.1.2). Nur durch die Konformationsänderung und die darauf sensitiv reagierende Lon-Protease kann das System dieses Verhalten zeigen. Diese Protease ist ebenfalls notwendig für die Bistabilität von RovA in *Y. pseudotuberculosis*.

Durch das Erstellen der *in silico*-Mutante  $\Delta$ iRovA konnte gezeigt werden, dass die NFL innerhalb des RovA-Systems nur einen Einfluss auf die Konzentration der Gleichgewichtszustände von niedrigen Temperaturen ausgehend hat. Die Konzentrationen sind bei Abwesenheit der NFL um ein vielfaches größer als mit NFL. Eine zu hohe Konzentration bei niedrigen Temperaturen kann auch die Anpassungszeit und den Energiekostenaufwand für *Y. pseudotuberculosis* wesentlich vergrößern. Den größten Einfluss auf die Bistabilität hat innerhalb dieses Modells die PFL und die Degradationsrate, welche beide notwendige Bedingungen für ein derartiges temperaturbe-

dingtes Verhalten darstellen. In wieweit die Vorhersagen der Wirklichkeit entsprechen, muss im Experiment geklärt werden.

## 11.6. Ausblick

In der mathematischen Modellierung der RovA-Produktion wurden verschiedene Experimente zu einem gesamten Modell vereint. Somit wurde ein gutes Beispiel für die Möglichkeiten der Zusammenarbeit der Biomathematik und der Systembiologie mit der experimentellen Biologie geliefert. Mit dieser Arbeit wurde ein Modell das bistabile Verhalten von RovA beschreibt dargestellt. Das Modell wurde genauer untersucht.

Da die Produktion von RovA ein temperaturabhängiges bistabiles Verhalten hat, bietet dieser Regulator die Möglichkeit der biotechnologischen Nutzung. Die RovA-Regulation kann Grundlage für einen genregulatorischen Schalter sein, der durch unterschiedliche Temperaturen an- beziehungsweise ausgeschaltet werden kann. Um dieses verwirklichen zu können, war das bistabile Verhalten zu erfassen und die Einflussfaktoren auf die Bistabilität zu identifizieren.



# Literaturverzeichnis

ACHBERGEROVÁ, L. AND NAHÁLKA, J. 2011. Polyphosphate – an ancient energy source and active metabolic regulator. *Microb. Cell. Fact.* 10:63.

ALON, U. 2007. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8:450–461.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W., AND LIPMAN, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.

ANGELI, D., FERRELL, JR., J. E., AND SONTAG, E. D. 2004. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. U.S.A.* 101:1822–1827.

ARNOLD, C. AND HODGSON, I. J. 1991. Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl.* 1:39–42.

ATTUR, M. G., DAVE, M. N., TSUNOYAMA, K., AKAMATSU, M., KOBORI, M., MIKI, J., ABRAMSON, S. B., KATOH, M., AND AMIN, A. R. 2002. A system biology approach to bioinformatics and functional genomics in complex human diseases: arthritis. *Curr. Issues Mol. Biol.* 4:129–146.

BECSKEI, A., SÉRAPHIN, B., AND SERRANO, L. 2001. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* 20:2528–2535.

BEDNARZ, M., HALLIDAY, J. A., HERMAN, C., AND GOLDING, I. 2014. Revisiting bistability in the lysis/lysogeny circuit of bacteriophage lambda. *PLoS One* 9:100876.

BIASINI, M., BIENERT, S., WATERHOUSE, A., ARNOLD, K., STUDER, G., SCHMIDT, T., KIEFER, F., CASSARINO, T., BERTONI, M., BORDOLI, L., AND SCHWEDE, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* .

BICKLE, T. A. 2004. Restricting restriction. *Mol. Microbiol.* 51:3–5.

BICKLE, T. A. AND KRÜGER, D. H. 1993. Biology of DNA restriction. *Microbiol. Rev.* 57:434–450.

BOTTONE, E. J. 1997. *Yersinia enterocolitica*: the charisma continues. *Clin. Microbiol. Rev.* 10:257–276.

BOTTONE, E. J. 1999. *Yersinia enterocolitica*: overview and epidemiologic correlates. *Microbes Infect.* 1:323–333.

BRUGGEMAN, F. J., HORNBERG, J. J., BOOGERD, F. C., AND WESTERHOFF, H. V. 2007. Introduction to systems biology. *EXS* 97:1–19.

BÖHME, K., STEINMANN, R., KORTMANN, J., SEEKIRCHER, S., HEROVEN, A. K., BERGER, E., PISANO, F., THIERMANN, T., WOLF-WATZ, H., NARBERHAUS, F., AND DERSCH, P. 2012. Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control *Yersinia* virulence. *PLoS Pathog.* 8:e1002518.

CHAN, P.-L., YU, V., WAI, L., AND YU, H.-F. 2006. Production of medium-chain-length polyhydroxyalkanoates by *Pseudomonas aeruginosa* with fatty acids and alternative carbon sources. *Appl. Biochem. Biotechnol.* 129-132:933–941.

- CHAUDHURI, R. R., PETERS, S. E., PLEASANCE, S. J., NORTHEN, H., WILLERS, C., PATERSON, G. K., CONE, D. B., ALLEN, A. G., OWEN, P. J., SHALOM, G., STEKEL, D. J., CHARLES, I. G., AND MASKELL, D. J. 2009. Comprehensive identification of *Salmonella enterica* serovar typhimurium genes required for infection of BALB/c mice. *PLoS Pathog.* 5:e1000529.
- CLEWELL, D. B. 1981. Plasmids, drug resistance, and gene transfer in the genus *Streptococcus*. *Microbiol Rev.* 45:409–436.
- COOPER, P. S., LIPSHULTZ, D., MATTEN, W. T., MCGINNIS, S. D., PECHOUS, S., ROMITI, M. L., TAO, T., VALJAVEC-GRATIAN, M., AND SAYERS, E. W. 2010. Education resources of the national center for biotechnology information. *Brief Bioinform.* 11:563–569.
- DEICHMANN, U., SCHUSTER, S., MAZAT, J.-P., AND CORNISH-BOWDEN, A. 2014. Commemorating the 1913 michaelis-menten paper die kinetik der invertin-wirkung: three perspectives. *FEBS J.* 281:435–463.
- DUBE, P. H., HANDLEY, S. A., REVELL, P. A., AND MILLER, V. L. 2003. The rovA mutant of yersinia enterocolitica displays differential degrees of virulence depending on the route of infection. *Infect. Immun.* 71:3512–3520.
- DUBNAU, D. AND LOSICK, R. 2006. Bistability in bacteria. *Mol. Microbiol.* 61:564–572.
- EICHENBERGER, P., FUJITA, M., JENSEN, S. T., CONLON, E. M., RUDNER, D. Z., WANG, S. T., FERGUSON, C., HAGA, K., SATO, T., LIU, J. S., AND LOSICK, R. 2004. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol.* 2:e328.
- FREY, U. H., BACHMANN, H. S., PETERS, J., AND SIFFERT, W. 2008. PCR-amplification of gc-rich regions: 'slowdown PCR'. *Nat. Protoc.* 3:1312–1317.

- GOODSELL, D. S. 2002. The molecular perspective: restriction endonucleases. *Stem Cells* 20:190–191.
- GRENET, E. AND ROBELIN, J. 1988. Analysis of the distribution of particle sizes at different levels of the digestive tract. *Reprod. Nutr. Dev.* 28 Suppl 1:139–140.
- GÖRKE, B. AND STÜLKE, J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiology* 6:613–624.
- HAHN, M. W. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* 100:605–617.
- HARTMAN, P. 1960. A lemma in the theory of structural stability of differential equations. *Proc. Amer. Math. Soc.* 11:610–620.
- HASTY, J., McMILLEN, D., AND COLLINS, J. J. 2002. Engineered gene circuits. *Nature* 420:224–230.
- HAYES, F. 2003. Transposon-based strategies for microbial functional genomics and proteomics. *Annu. Rev. Genet.* 37:3–29.
- HENSEL, M., SHEA, J. E., GLEESON, C., JONES, M. D., DALTON, E., AND HOLDEN, D. W. 1995. Simultaneous identification of bacterial virulence genes by negative selection. *Science* 269:400–403.
- HERBST, K., BUJARA, M., HEROVEN, A. K., OPITZ, W., WEICHERT, M., ZIMMERMANN, A., AND DERSCH, P. 2009. Intrinsic thermal sensing controls proteolysis of *Yersinia* virulence regulator RovA. *PLoS Pathog.* 5:e1000435.
- HEROVEN, A. K., NAGEL, G., TRAN, H. J., PARR, S., AND DERSCH, P. 2004. RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and RovA expression in *Yersinia pseudotuberculosis*. *Mol. Microbiol.* 53:871–888.



- HEROVEN, A. K., SEST, M., PISANO, F., SCHEB-WETZEL, M., STEINMANN, R., BÖHME, K., KLEIN, J., MÜNCH, R., SCHOMBURG, D., AND DERSCH, P. 2012. Crp induces switching of the CsrB and CsrC RNAs in *Yersinia pseudotuberculosis* and links nutritional status to virulence. *Front Cell Infect. Microbiol.* 2:158.
- HILL, C. W. 1999. Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. *Res. Microbiol.* 150:665–674.
- HOLLOWAY, B. W. 1969. Genetics of *Pseudomonas*. *Bacteriol. Rev.* 33:419–443.
- HORNUNG, G. AND BARKAI, N. 2008. Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. *PLoS Comput. Biol.* 4:e8.
- IDEKER, T., GALITSKI, T., AND HOOD, L. 2001. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* 2:343–372.
- JUDSON, N. AND MEKALANOS, J. J. 2000. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nat. Biotechnol.* 18:740–745.
- JUNGCK, J. R. AND WEISSTEIN, A. E. 2013. Mathematics and evolutionary biology make bioinformatics education comprehensible. *Brief Bioinform.* 14:599–609.
- KALIR, S., MANGAN, S., AND ALON, U. 2005. A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol. Syst. Biol.* 1:2005.0006.
- KANEHISA, M., FICKETT, J. W., AND GOAD, W. B. 1984. A relational database system for the maintenance and verification of the los alamos sequence library. *Nucleic Acids Res.* 12:149–158.

- KONKEL, M. E. AND TILLY, K. 2000. Temperature-regulated expression of bacterial virulence genes. *Microbes Infect.* 2:157–166.
- KOORNHOF, H. J., SMEGO, JR, R., AND NICOL, M. 1999. Yersiniosis. ii: The pathogenesis of *Yersinia* infections. *Eur. J. Clin. Microbiol. Infect. Dis.* 18:87–112.
- LAFAYE DE MICHEAUX, P., DROUILHET, R., AND LIQUET, B. 2011. The R software. Mastering the language. Effecting statisticac analyses. (Le logiciel R. Maîtriser le langage. Effectuer des analyses statistiques.). Paris: Springer.
- LANGEMANN, D. AND REHBERG, M. 2010. Unbuffered and buffered supply chains in human metabolism. *J. Biol. Phys.* 36:227–244.
- LARGAESPADA, D. A. 2009. Transposon mutagenesis in mice. *Methods Mol. Biol.* 530:379–390.
- LATHAM, P. E., RICHMOND, B. J., NELSON, P. G., AND NIRENBERG, S. 2000. Intrinsic dynamics in neuronal networks. i. theory. *J. Neurophysiol.* 83:808–827.
- LIELEG, O., CALDARA, M., BAUMGÄRTEL, R., AND RIBBECK, K. 2011. Mechanical robustness of *Pseudomonas aeruginosa* biofilms. *Soft Matter* 7:3307–3314.
- LIPMAN, D. J. AND PEARSON, W. R. 1985. Rapid and sensitive protein similarity searches. *Science* 227:1435–1441.
- LOTKA, A. J. 1922. Contribution to the energetics of evolution. *Proc. Natl. Acad. Sci. U.S.A.* 8:147–151.
- LURIA, S. E. AND HUMAN, M. L. 1952. A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.* 64:557–569.

- LYCZAK, J. B., CANNON, C. L., AND PIER, G. B. 2002. Lung infections associated with cystic fibrosis., volume 15. *Clin. Microbiol. Rev.*
- MA, J., WANG, P. W., YAO, D., WANG, Y. P., YAN, W., AND GUAN, S. C. 2011. Single-primer PCR correction: a strategy for false-positive exclusion. *Genet. Mol. Res.* 10:150–159.
- MAAMAR, H. AND DUBNAU, D. 2005. Bistability in the *Bacillus subtilis* k-state (competence) system requires a positive feedback loop. *Mol. Microbiol.* 56:615–624.
- MACIA, J., WIDDER, S., AND SOLÉ, R. 2009. Specialized or flexible feed-forward loop motifs: a question of topology. *BMC Syst. Biol.* 3:84.
- MANGAN, S. AND ALON, U. 2003. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U.S.A.* 100:11980–11985.
- MANGAN, S., ITZKOVITZ, S., ZASLAVER, A., AND ALON, U. 2006. The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J. Mol. Biol.* 356:1073–1081.
- MANGAN, S., ZASLAVER, A., AND ALON, U. 2003. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.* 334:197–204.
- McCLINTOCK, B. 1950. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.* 36:344–355.
- McCLINTOCK, B. 1953. Induction of instability at selected loci in maize. *Genetics* 38:579–599.
- MITROPHANOV, A. Y. AND GROISMAN, E. A. 2008. Positive feedback in cellular control systems. *Bioessays* 30:542–555.

MORITA, Y., TOMIDA, J., AND KAWAMURA, Y. 2014. Responses of *Pseudomonas aeruginosa* to antimicrobials. *Front. Microbiol.* 4:422.

NAGEL, G., LAHRZ, A., AND DERSCH, P. 2001. Environmental control of invasion expression in *Yersinia pseudotuberculosis* is mediated by regulation of RovA, a transcriptional activator of the SlyA/Hor family. *Mol. Microbiol.* 41:1249–1269.

NARTEN, M., ROSIN, N., SCHOBERT, M., AND TIELEN, P. 2012. Susceptibility of *Pseudomonas aeruginosa* urinary tract isolates and influence of urinary tract conditions on antibiotic tolerance. *Curr. Microbiol.* 64:7–16.

NEEDLEMAN, S. B. AND WUNSCH, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443–453.

NOMENCLATURE, C. B. 1970. Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. recommendations 1970. *Biochem. J.* 120:449–454.

OCHMAN, H., GERBER, A. S., AND HARTL, D. L. 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics* 120:621–623.

OEHLER, S., EISMANN, E. R., KRÄMER, H., AND MÜLLER-HILL, B. 1990. The three operators of the lac operon cooperate in repression. *EMBO J.* 9:973–979.

OLIVA, W. M. AND SALLUM, E. M. 1996. Periodic dynamic systems for infected hosts and mosquitoes. *Rev. Saude Publica* 30:218–223.

OPPERMAN, T., LING, L. L., AND MOIR, D. T. 2003. Microbial pathogen genomes - new strategies for identifying therapeutic and vaccine targets. *Expert Opin Ther Targets* 7:469–473.

- OSTERMAN, I. A., EVFRATOV, S. A., SERGIEV, P. V., AND DONTSOVA, O. A. 2013. Comparison of mRNA features affecting translation initiation and reinitiation. *Nucleic Acids Res.* 41:474–486.
- OZBUDAK, E. M., THATTAI, M., LIM, H. N., SHRAIMAN, B. I., AND OUDENAARDEN, A. V. 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427:737–740.
- PICARDEAU, M. 2010. Transposition of fly mariner elements into bacteria as a genetic tool for mutagenesis. *Genetica* 138:551–558.
- PINGOUD, A., FUXREITER, M., PINGOUD, V., AND WENDE, W. 2005. Type ii restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.* 62:685–707.
- POMERENING, J. R., SONTAG, E. D., AND FERRELL, JR, J. E. 2003. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nat. Cell Biol.* 5:346–351.
- QUADE, N., MENDONCA, C., HERBST, K., HEROVEN, A. K., RITTER, C., HEINZ, D. W., AND DERSCH, P. 2012. Structural basis for intrinsic thermosensing by the master virulence regulator RovA of *Yersinia*. *J. Biol. Chem.* 287:35796–35803.
- REINKE, J. 1901. Einleitung in Die Theoretische Biologie. Nabu Press.
- REVELL, P. A. AND MILLER, V. L. 2000. A chromosomally encoded regulator is required for expression of the *Yersinia enterocolitica* *inv* gene and for virulence. *Mol. Microbiol.* 35:677–685.
- ROMERO, D., MARTÍNEZ-SALAZAR, J., ORTIZ, E., RODRÍGUEZ, C., AND VALENCIA-MORALES, E. 1999. Repeated sequences in bacterial chromosomes and plasmids: a glimpse from sequenced genomes. *Res. Microbiol.* 150:735–743.

ROSELIUS, L., LANGEMANN, D., MÜLLER, J., HENSE, B. A., FILGES, S., JAHN, D., AND MÜNCH, R. 2014. Modelling and analysis of a gene-regulatory feed-forward loop with basal expression of the second regulator. *J Theor Biol* 363C:290–299.

SANGER, F. AND TUPPY, H. 1951. The amino-acid sequence in the phenylalanyl chain of insulin. 2. the investigation of peptides from enzymic hydrolysates. *Biochem. J.* 49:481–490.

SANMIGUEL, P. AND BENNETZEN, J. L. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany*. 82:37–44.

SAYERS, E. W., BARRETT, T., BENSON, D. A., BOLTON, E., BRYANT, S. H., CANESE, K., CHETVERNIN, V., CHURCH, D. M., DICUCCIO, M., FEDERHEN, S., FEOLO, M., FINGERMAN, I. M., GEER, L. Y., HELMBERG, W., KAPUSTIN, Y., LANDSMAN, D., LIPMAN, D. J., LU, Z., MADDEN, T. L., MADEJ, T., MAGLOTT, D. R., MARCHLER-BAUER, A., MILLER, V., MIZRACHI, I., OSTELL, J., PANCHENKO, A., PHAN, L., PRUITT, K. D., SCHULER, G. D., SEQUEIRA, E., SHERRY, S. T., SHUMWAY, M., SIROTKIN, K., SLOTTA, D., SOUVOROV, A., STARCHENKO, G., TATUSOVA, T. A., WAGNER, L., WANG, Y., WILBUR, W. J., YASCHENKO, E., AND YE, J. 2011. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 39:D38–D51.

SCHNEIDER, C. A., RASBAND, W. S., AND ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9:671–675.

SCHNEIDER, M. V., WATSON, J., ATTWOOD, T., ROTHER, K., BUDD, A., MCDOWALL, J., VIA, A., FERNANDES, P., NYRONEN, T., BLICHER, T., JONES, P., BLATTER, M.-C., DE LAS RIVAS, J., JUDGE, D. P., VAN DER GOOL, W., AND BROOKSBANK, C. 2010. Bioinformatics training: a review of challenges, actions and support requirements. *Brief Bioinform.* 11:544–551.

- SCHOBERT, M. AND JAHN, D. 2010. Anaerobic physiology of *Pseudomonas aeruginosa* in the cystic fibrosis lung. *Int. J. Med. Microbiol.* 300:549–556.
- SCHOBERT, M. AND TIELEN, P. 2010. Contribution of oxygen-limiting conditions to persistent infection of *Pseudomonas aeruginosa*. *Future Microbiol.* 5:603–621.
- SCHREIBER, K., BOES, N., ESCHBACH, M., JAENSCH, L., WEHLAND, J., BJARNSHOLT, T., GIVSKOV, M., HENTZER, M., AND SCHOBERT, M. 2006. Anaerobic survival of *Pseudomonas aeruginosa* by pyruvate fermentation requires an Usp-type stress protein. *J. Bacteriol.* 188:659–668.
- SCHREIBER, K., KRIEGER, R., BENKERT, B., ESCHBACH, M., ARAI, H., SCHOBERT, M., AND JAHN, D. 2007. The anaerobic regulatory network required for *Pseudomonas aeruginosa* nitrate respiration. *J. Bacteriol.* 189:4310–4314.
- SHEN-ORR, S. S., MILO, R., MANGAN, S., AND ALON, U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31:64–68.
- SHEVCHUK, O., BATZILLA, C., HÄGELE, S., KUSCH, H., ENGELMANN, S., HECKER, M., HAAS, A., HEUNER, K., GLÖCKNER, G., AND STEINERT, M. 2009. Proteomic analysis of *Legionella*-containing phagosomes isolated from dictyostelium. *Int. J. Med. Microbiol.* 299:489–508.
- SHEVCHUK, O., ROSELIUS, L., GÜNTHER, G., KLEIN, J., JAHN, D., STEINERT, M., AND MÜNCH, R. 2012. InFiRe – a novel computational method for the identification of insertion sites in transposon mutagenized bacterial genomes. *Bioinformatics* 28:306–310.
- SHEVCHUK, O. AND STEINERT, M. 2009. Screening of virulence traits in *Legionella pneumophila* and analysis of the host susceptibility to infection by using the Dictyostelium host model system. *Methods Mol. Biol.* 470:47–56.

SMITH, V., BOTSTEIN, D., AND BROWN, P. O. 1995. Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc. Natl. Acad. Sci. U.S.A.* 92:6479–6483.

SNOEP, J. L., BRUGGEMAN, F., OLIVIER, B. G., AND WESTERHOFF, H. V. 2006. Towards building the silicon cell: a modular approach. *Biosystems* 83:207–216.

SOUTHERN, E. 2006. Southern blotting. *Nat. Protoc.* 1:518–525.

SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.

STEKEL, D. J. AND JENKINS, D. J. 2008. Strong negative self regulation of prokaryotic transcription factors increases the intrinsic noise of protein expression. *BMC Syst. Biol.* 2:6.

STEUER, R., GROSS, T., SELBIG, J., AND BLASIUS, B. 2006. Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 103:11868–11873.

STROHM, T. O., GRIFFIN, B., ZUMFT, W. G., AND SCHINK, B. 2007. Growth yields in bacterial denitrification and nitrate ammonification. *Appl. Environ. Microbiol.* 73:1420–1424.

TIELEN, P., ROSIN, N., MEYER, A.-K., DOHNT, K., HADDAD, I., JÄNSCH, L., KLEIN, J., NARTEN, M., POMMERENKE, C., SCHEER, M., SCHOBERT, M., SCHOMBURG, D., THIELEN, B., AND JAHN, D. 2013. Regulatory and metabolic networks for the adaptation of *Pseudomonas aeruginosa* biofilms to urinary tract-like conditions. *PLoS One* 8:71845.

TRAN, H. J., HEROVEN, A. K., WINKLER, L., SPRETER, T., BEATRIX, B., AND DERSCH, P. 2005. Analysis of RovA, a transcriptional regulator of *Yersi-*



*nia pseudotuberculosis* virulence that acts through antirepression and direct transcriptional activation. *J. Biol. Chem.* 280:42423–42432.

TRUNK, K., BENKERT, B., QUÄCK, N., MÜNCH, R., SCHEER, M., GARBE, J., JÄNSCH, L., TROST, M., WEHLAND, J., BUER, J., JAHN, M., SCHOBERT, M., AND JAHN, D. 2010. Anaerobic adaptation in *Pseudomonas aeruginosa*: definition of the anr and dnr regulons. *Environ. Microbiol.* 12:1719–1733.

TSIEN, R. Y. 1998. The green fluorescent protein. *Annu. Rev. Biochem.* 67:509–544.

TYLER, B., LOOMIS, JR, W., AND MAGASANIK, B. 1967. Transient repression of the lac operon. *J. Bacteriol.* 94:2001–2011.

VEENING, J.-W., HAMOEN, L. W., AND KUIPERS, O. P. 2005. Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. *Mol. Microbiol.* 56:1481–1494.

VILEN, H., AALTO, J.-M., KASSINEN, A., PAULIN, L., AND SAVILAHTI, H. 2003. A direct transposon insertion tool for modification and functional analysis of viral genomes. *J. Virol.* 77:123–134.

VILLANI, M. AND SERRA, R. 2013. On the dynamical properties of a model of cell differentiation. *EURASIP J. Bioinform. Syst. Biol.* 2013:4.

VOLTERRA, V. 1931. Variations and fluctuations of the number of individuals in animal species living together in Animal Ecology. Chapman, R.N. (ed), McGraw–Hill.

WALL, M. E., DUNLOP, M. J., AND HLAVACEK, W. S. 2005. Multiple functions of a feed-forward-loop gene circuit. *J. Mol. Biol.* 349:501–514.

WETTENHALL, J. M. AND SMYTH, G. K. 2004. limmaGUI: A graphical user interface for linear modeling of microarray data. *Bioinformatics* .

WICKER, T., SABOT, F., HUA-VAN, A., BENNETZEN, J. L., CAPY, P., CHALHOUB, B., FLAVELL, A., LEROY, P., MORGANTE, M., PANAUD, O., PAUX, E., SANMIGUEL, P., AND SCHULMAN, A. H. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973–982.

WIDDER, S., SOLÉ, R., AND MACÍA, J. 2012. Evolvability of feed-forward loop architecture biases its abundance in transcription networks. *BMC Syst. Biol.* 6:7.

WOLFSBERG, T. G. 2011. Using the NCBI Map Viewer to browse genomic sequence data. *Curr. Protoc. Hum. Genet.* Chapter 18:Unit18.5.

XIONG, W. AND FERRELL, J. E. 2003. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* 426:460–465.

XU, K. D., STEWART, P. S., XIA, F., HUANG, C. T., AND McFETERS, G. A. 1998. Spatial physiological heterogeneity in *Pseudomonas aeruginosa* biofilm is determined by oxygen availability. *Appl. Environ. Microbiol.* 64:4035–4039.

YAGIL, G. AND YAGIL, E. 1971. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* 11:11–27.

YERGEAU, D. A. AND MEAD, P. E. 2009. Transposon-mediated transgenesis in the frog: New tools for biomedical and developmental studies. *Front. Biosci.* 14:225–236.

YOON, S. S., KARABULUT, A. C., LIPSCOMB, J. D., HENNIGAN, R. F., LYMAR, S. V., GROCE, S. L., HERR, A. B., HOWELL, M. L., KILEY, P. J., SCHURR, M. J., GASTON, B., CHOI, K.-H., SCHWEIZER, H. P., AND HASSETT, D. J. 2007. Two-pronged survival strategy for the major cystic fibrosis pathogen, *Pseudomonas aeruginosa*, lacking the capacity to degrade nitric oxide during anaerobic respiration. *EMBO J.* 26:3662–3672.

- 
- ZUMFT, W. G. 1997. Cell biology and molecular basis of denitrification. *Microbiol Mol. Biol. Rev.* 61:533–616.



# Index

Überlappung, 56

Abbaurate, 168

Ableitung der Nutzenbilanz-Funktion,  
110

Ableseungenauigkeit, 47, 51, 56

Aktivierung, 21

Atmung, 75

Aufbau, 140, 142

Autoregulation, 154, 159

Bakterienarten

*Bacillus subtilis*, 22, 159

*Escherichia coli*, 21, 84, 158

*Legionella pneumophila*, 31, 40,  
64

*Pseudomonas aeruginosa*, 74, 143,  
145

*Pseudomonas putida*, 22

*Yersinia enterocolitica*, 149

*Yersinia pestis*, 149

*Yersinia pseudotuberculosis*, 151,  
159, 177

*Yersinia pseudotuberculosis*, 149,

150

Basale Expression, 84

Bindekonstanten, 170

Bioconductor, 16

Biofilm, 74, 75

Bistabilität, 149, 158, 159

Blunt ends, 35

Chloramphenicol, 168

Dateiformate

Fasta-Format, 17, 61

gbk-Format, 17

Rd-Format, 17

Denitrifikation, 75

Dichtefunktion, 39, 40

Differentialgleichung, 23, 87

Digestion

Paket, 48

Dimerisierung, 154

Dissimilatorische Nitratreduktase, 79

DNA Data Bank of Japan (DDBJ),

11

- EcoRI, 40
- Energie, 84
- Energiespeicher, 119, 122, 127
- Energieverteiler, 119, 122, 127
- Enteropathogen, 150
- European Molecular Biology Laboratory (EMBL), 11
- European Nucleotide Archive (ENA), 11
- Evolutionärer Vorteil, 107, 116, 143
- Exponentialverteilung, 39
- FBL, feedback loop, 22, 153, 156
- Fehlerwahrscheinlichkeit, 55, 56
- FFL, feedforward loop, 21, 80
- Filterung, 82
- Fitness, 121, 123, 134
- Fragmentlängenverteilung, 38
- Gedächtnis, 105, 141
- Gelelektrophorese, 36
- Genbank, 11
- Genregulation, 18
- Genregulatorische Netzwerke, 20
- Gleichgewichtszustand, 141
- Grenzbereich, 113, 123
- Grundmetabolismus, 121
- Halbsättigungskonstante, 108
- Harnwegsinfektionen, 74
- Hartman-Grobman-Theorem, 100
- Hill-Funktion, 25
- Hill-Koeffizient, 25, 108
- Hysteresis, 149, 158, 161
- Infektion, 150
- InFiRe
  - Algorithmus, 43
  - Arbeitsablauf, 45
  - Paket, 49
  - Statistik, 50
    - Explizite Statistik, 54
    - Globale Statistik, 51
    - Vergleich, 59
  - Webinterface, 60
    - Analyse, 60
    - Explizite Statistik, 58
- Inkubationszeit, 150
- INSDC (International Nucleotide Sequence Database Collaboration), 10
- Izumi-Fieber, 150
- Kostenfaktor, 112
- Landau-Symbol, 99
- Lineares Modell, 98
- Lineares Modell mit Speicher, 124
- Linearisierungssatz, 100
- Lon-Protease, 154
- Los Alamos National Laboratory (LANL), 10
- Matlab, 15
- Michaelis-Menten-Konstante, 25

- Modell mit Speicher, 118  
Modellierung, 86, 164  
Mukus, 75  
  
NAP, 79  
NCBI, 37, 61, 63  
NCBI map viewer, 63  
NFL, negative feedback loop, 156  
Nichtlineares Modell, 107  
Nichtlineares Modell mit Speicher, 128  
Nitratatmung, 79  
    Anr, 77, 79  
    Antiporter, 79  
    Cytochrom cd1, 79  
    Distickstoffmonoxid-Reduktase, 75, 79, 80  
    Dnr, 79  
    Häm d1, 79  
    NAR, 79  
    NarK, 79  
    NarL, 79  
    NarLP, 79  
    NarX, 79  
    NarXL, 77, 79, 84  
    NIR, 79  
    NirO, 79  
    NirP, 79  
    NirQ, 79  
    NirS, 79  
    Nitrat-Nitrit-Antiporter, 79  
    Nitratreduktase, 75, 79  
    Nitratsensorsystem, 79  
    Nitritreduktase, 75, 79, 80  
    Nor, 79  
    NOS, 79  
    NosL, 79  
    NosR, 79  
    Stickstoffmonoxid-Reduktase, 75, 79, 80  
Normierung, 142  
Notwendige Bedingung, 136  
Nutzenbilanz, 105  
Nylonmembran, 36  
  
Operator L, 95, 102, 110  
Optimale basale Expression, 140  
Oxidation, 75  
  
PCR, 48, 63  
Person  
    Moler, Cleve, 15  
    Goad, Walter, 11  
    Klein, Johannes, 178  
    Müller, Johannes, 178  
    McClintock, Barbara, 32  
    Nuss, Aron, 178  
    Pfeffer, Claude, 10  
    Southern, Edwin, 36  
PFL, positive feedback loop, 157, 159  
Phasendiagramme, 162  
Primer-BLAST, 64

- 
- Produktionsfunktionen, 98, 108
- Promotor, 155
- Promotorbindung, 170
- Pseudocode, 52
- R, 16
- R-Pakete, 48
- Reaktionsgeschwindigkeit, 23
- Regulator, 21, 80
- Reprimierung, 21
- Responseregulator, 77
- Restriktionsenzym, 34, 36
- Restriktionsreaktion, 56
- RovA, 149, 151, 152
- RovA-Modell, 180
- Sättigungsfunktionen, 108
- Sensorkinase, 77
- Signal, 87, 99
- Simulation, 104, 116, 125, 132
- Southern Blot, 36
- Stäbchen, 74
- Stickstoff, 76
- Sticky ends, 35
- Strenges Maximum, 140
- Substrat, 25
- t-Test, 60
- Temperatur, 153
- Temperaturwechsel Experiment, 176
- Transposon, 32
- Klasse1, 33
- Klasse2, 33
- Mutagenese, 33
- Retrotransposon, 33
- Transposon-Klassen, 32
- Vergleich der Modelle, 139
- Virulenz, 151
- Virulenzregulator, 149
- Wachstum, 121
- Webinterface, 48, 61
- Zeitverzögerung, 143
- Zielvariable, 80
- Zwei-Komponenten-System, 20
- zystische Fibrose, 75





# Danksagung

Ich möchte Prof. Dr. Dieter Jahn danken, der mir in den letzten Jahren die Möglichkeit gegeben hat zu promovieren. Vielen Dank für die Unterstützung in der Zeit und die guten Ratschläge. Vielen Dank für die tolle Zeit. Es ist nicht selbstverständlich, einen so netten Chef zu haben.

Ebenfalls möchte ich Prof. Dr. Dirk Langemann danken, mit dem ich viele Stunden der Diskussion verbracht habe. Vielen Dank für die guten Ratschläge, Unterstützungen und Vorschläge. Ebenfalls möchte ich mich bei ihm für die Bereitschaft, das Zweitgutachten dieser Dissertation anzufertigen bedanken. Er hat mir immer sehr mit meiner Arbeit geholfen.

Prof. Dr. Michael Steinert möchte ich sowohl für Kooperationen als auch für die Übernahme des Vorsitzes der Prüfungskommission danken.

Ein ganz besonderer Dank gilt Dr. Richard Münch, der mich immer unterstützt hat und den ich immer fragen durfte. Richard Münch war nicht nur beruflich für mich da, sondern hat mich auch privat immer unterstützt. Ich danke ihm sehr für seine große Hilfe.

Dr. Olga Shevchuk möchte ich für die gute Zusammenarbeit danken. Sie ist für mich ein Vorbild.

Für die Korrektur möchte ich Steffi Heyber danken. Sie ist nicht nur eine tolle Kollegin für mich, sondern auch eine ganz besondere Freundin geworden.

Ebenfalls möchte ich mich bei vielen, weiteren Menschen für Kooperationen, Hilfen und Unterstützung bedanken: Prof. Dr. Johannes Müller, Prof. Dr. Petra Dersch, PD Dr. Max Schobert, Dr. Katja Böhme, Dr. Burkhard Hense, Dr. Johannes Klein, Dr. Aaron Nuss, Stefan Filges, Jenny Jacobs, Karin Münch und Sarah Wienecke.

Danke auch an das gesamte Institut für Mikrobiologie und speziell an die Arbeitsgruppe Münch. Es war eine tolle Zeit, die ich nie vergessen werde.

Ich möchte mich beim Bundesministerium für Bildung und Forschung (BMBF) und bei der deutschen Forschungsgesellschaft (DFG) für die Bereitstellung der Finanzierung meiner Promotion bedanken.

Ein besonders großer Dank geht an meine Familie, meinen Ehemann Daniel Thiele, meinen Eltern Maren und Dieter Roselius und meiner Schwester Catina Roselius. Danke für Feedback und Korrektur. Sie waren immer für mich da und haben mich immer unterstützt.

## Dezimalzeichen

Aus Gründen der Übersichtlichkeit wurde als Dezimaltrennzeichen ein Punkt verwendet.



